EOSC Future

EOSC in Practice Stories

Acknowledgements

Disclaimer

Table of Contents

Introduction

#EOSCSYMPOSIUM23

The European Science Cloud EOSC vision is to set up a "Web of FAIR Data and Services" for science and scholarship in Europe. By harnessing the potential of Open Science to meet the need for collaborative science to address the societal challenges we face, EOSC can transform how data-driven science is realised by providing European researchers, innovators, companies and citizens access to research data, research infrastructures, einfrastructures and related services. EOSC can empower users to publish, find and reuse data, tools and services for research, innovation and educational purposes.

The EOSC-Future project has played an important role in the implementation phase of EOSC, integrating and consolidating existing data and services from science communities, research infrastructures and e-infrastructures. The project has also been a catalyst for bringing together key players in the EOSC community, collecting different perspectives, requirements and achievements to ensure that all voices are heard as EOSC evolves.

As we look ahead, the EOSC Multi-Annual Roadmap cites the importance of engagement and the widening of EOSC through the expansion of use cases. This collection of over 30 EOSC in Practice Stories, gathers a variety of insights and examples of how EOSC is steadily consolidating a narrative of key users and providers citing the benefits of EOSC in researchers' work . Experiences from various disciplines highlight how EOSC can strive to ensure scientists can get the maximum out of research data through access to complementary scientific and analytical services from multiple sources across Europe.

The stories provide real examples of how existing data- and e-infrastructures which were previously dispersed across disciplines and the EU Member States, are being federated with the key goal of being established around a core which follows frameworks, principles and rules. The stories give practical examples of this, addressing challenges around data sovereignty, standards and regulations, while ensuring the data are findable, accessible, interoperable and reusable - FAIR.

Gathered over the last 24 months, the stories also reflect the transition between the European Commission funding programmes Horizon 2020 and Horizon Europe and the move from the EOSC Prototype to roll out phases.

Where we start?

Our stories start from an implementation perspective showcasing how the EOSC Future project has provided a broad range of resources and services for the full lifecycle of research data which aim to help break open disciplinary silos. Addressing different scientific challenges and domains, the "Science Projects" demonstrate the integration of research data and services across scientific domains. They also highlight the role EOSC can play in each stage of the research process. The Science Projects also monitor researchers' real-life data needs and requirements for cross-domain and composability features.

EOSC Future has also been instrumental in providing funding for different stakeholder groups to contribute to the roll out of EOSC. The RDA/EOSC Future open-call series has made real impact providing over 70 grants to the value of €1 million. We highlight a selection of these which show the opportunities created for scientific communities to contribute to the development of research data tools, standards and expertise in the context of EOSC. This has also brought data initiatives and experts closer to EOSC.

As part of an adoption funding programme supported by the EU and building on experience gained through the GÉANT-led OCRE project, the EOSC Future project allotted a total of €4.8 million across three calls, aimed at bridging the gap between commercial service providers and researchers through the EOSC ecosystem. We provide stories from three of the five projects which have made commercial services more accessible to Research & Education communities via the EOSC Marketplace, to digital service aggregators (e.g. non-profit entities, NRENs, RIs and e-Infrastructures, HPC centres, etc.) teaming up with OCRE cloud service providers.

EOSC Future also organised open procurement calls with the EOSC Digital Innovation Hub and we showcase how SMEs are now providing services through the EOSC Marketplace.

A key role of EOSC Future has also been to consolidate and build upon outputs of other key European Commission EOSC-related calls since 2018. These have also been captured in this document with stories focussing on different aspects of the evolution of different EOSC resources from the visualisation of complex and large data sets to virtual research environments. Thanks to EOSC, these are now available to a broader range of scientific communities, facilitating the sharing of materials and tools, the reproducibility and reuse of scientific methods.

This booklet has been produced for publication at the EOSC Symposium 2023 to showcase the valuable activities carried out by the scientific community and provide further evidence of the impact of EOSC on both providers and users of its resources and services.

More EOSC in Practice Stories are planned for publication as we move towards the end of the EOSC Future project in March 2024.

Access the full and continuously updated collection of EOSC in Practice Stories

Contribute with your own EOSC in Practice Story then visit the EOSC-Portal

Stories from the EOSC Future Project

Integrating research data and services across scientific domains

Stories from the EOSC Future Science Projects

The Dark Matter Science Project

The Dark Matter Test Science Project (TSP) within the European Science Cluster of Astronomy and Particle physics ESFRI research infrastructures (ESCAPE) project

The Dark Matter Science Project researchers:

- Ian Bird
- Fritjof Bengtsson;
- Pooja Bhattacharjee;
- Christopher Eckner;
- Per Alexander Ekman;
- Axel Gallen;
- Elena Gazzarrini;
- Tetiana Hryn'ova;
- Valerio Ippolito;
- Jared Little;
- Paolo Salomone;
- Mikhail Smirnov

The Dark Matter Science Project coordinators

Caterina Doglioni
caterina.doglioni@cern.ch

Lukas Heinrich
lukas.heinrich@cern.ch

Francesca Calore
francesca.calore@lapth.cnrs.fr

The Project Involved

ESCAPE
European Science Cluster of Astronomy &
Particle physics ESFRI research infrastructures

This project is a collaboration between scientists in European Research Infrastructures and experiments seeking to explain the nature of dark matter (such as HL-LHC, KM3NeT, CTA, DarkSide). The goal of this Science Project is to highlight the synergies between different dark matter communities and experiments, by producing new scientific results as well as by making the necessary data and software tools fully available.

The Research Community

Many of the European Research Infrastructures within ESCAPE have experiments that are searching for Dark Matter. There is a clear complementarity between these experiments under a variety of dark matter hypotheses.
Connecting results and potential discoveries from different experiments requires the engagement of all scientific communities involved – astrophysics, particle physics and nuclear physics – as already recommended within the update of the European Strategy of Particle Physics.

The Challenge

Besides the interpretation of results in terms of dark matter theories, synergies also exist between different communities and experiments in the tools needed to produce those results, in particular in terms of data management, data analysis and computing.
ESCAPE researchers funded by EOSC-Future have developed a Virtual Research Environment (VRE) that takes full advantage of these synergies. The VRE hosts both FAIR data and software, so that data analyses from different experiments can be executed in full.

EOSC service or tool used

The following lists the services developed by ESCAPE that are essential to this science environment and are contributed to EOSC Future:
- AAI: A fully developed AAI solution following the AARC blueprint, so scientists in the TSP's can use a single user identity for all aspects of work
- Data Lake: federated storage services with open and embargoed datasets
- Rucio – data location catalogue and policy engine.ESCAPE FTS (File Transfer Service) – for moving data around; integrated with the AAI service and the storage endpoints. Additionally: caching and streaming services to deliver data to processing and analysisA software catalogue to publish all of the needed analysis components, and make them available for the various groups involved in the TSP work
- Virtual Research Environment: An analysis environment, with a Jupyter notebook deployment, and access to scalable compute resources, tying in all these services.

Benefits and impact

- Entry point to the Virtual Research Environment where the workflows are implemented: https://escape2020.pages.in2p3.fr/virtual-environment/home/
- Delivered the services above, and implemented five different data analyses from experiments looking for dark matter (more being onboarded) on the VRE
- Talk & proceedings at the TOOLS20 conference (review of state-of-the-art tools for high energy physics and cosmology
- Engagement via Consortia (ECFA, APPEC, NUPECC, JENAS)
  - DM is a recognised challenge for the communities
  - Invited poster by DM postdoc at the JENAS conference 2022
  - Talk and upcoming paper from the FAIR4HEP workshop, to advance data & artificial intelligence (AI) models through the development of FAIR frameworks
  - Inputs to Snowmass white papers submitted to arXiv and journals [1][2][3][4] (Snowmass: input to prioritisation of particle physics, deciding US's science priorities). Figures made with code that is / is being onboarded to EOSC
- Two plenary talks at the biggest international conference for computing in high energy physics (CHEP), one on the DM Science Project and another one on the VRE [1][2]
- A full list of publications can be found at this link

Why do I need EOSC?

The Open Science added value for this Test Science Project is that all the digital objects within these new DM analyses will be implemented within the ESCAPE services infrastructure (Data Lake, Software Catalogue, Analysis Platform). We will make use of the ESCAPE Data Infrastructure for Open Science in the European Open Science Cloud to store, distribute and provide data and software access to the broad dark matter scientific community. This is a unique link between DM as a fundamental science question and the Open Science services needed to answer it that benefits the scientific community as a whole. This will serve as a stepping stone to build a virtual research environment within the European Open Science Cloud (EOSC).

Useful material related to this story

dark matter project
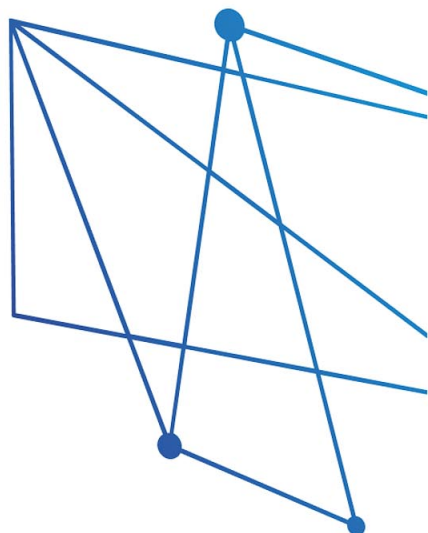
EOSC Future

Advancing Multi-disciplinary Research through Cross-Domain Integration and Collaboration

Enhancing Multi-disciplinary Research through Cross-Domain Integration and Collaboration

Hilde Orten

Norwegian Agency for Shared Services in Education and Research, Bergen, Norway

"By involving experts from the different partner organizations, and adopting standards aimed at cross-domain FAIR implementation, we can provide the additional visibility into the integration methods and processes to effectively address relevant multi-disciplinary research questions."

The Project Involved

SSHOC **ENVRI** community

EOSC Future Science Project 9, "Climate Neutral and Smart Cities" (H2020 Grant Agreement 101017536), focuses on multi-disciplinary research through collaboration between the SSHOC cluster and ENVRI-FAIR. Many experts and institutions are participating, including climate and air-quality scientists from the Swedish National Data Service, IAGOS, the Norwegian Center for Climate Service, and NILU, as well as experts on social survey data from the European Social Survey ERIC and CESSDA ERIC.

The User Community

This project has two main user communities: researchers engaged in multi-disciplinary projects and institutions dedicated to providing integrateable data effectively, both at scientific and technical levels.

The Challenge

Multiple levels of interoperability are considered, encompassing organizations, researchers, and required data resources and systems. The project supports collaborative research by integrating survey data from the European Social Survey, climate data from Copernicus, and air quality data from the European Environmental Agency.

This research examines environmental factors and their impact on urban inhabitants in various European cities, necessitating expertise across domains, access to diverse data sources, and agreement on semantics and technical implementation. The collaboration process is explored, and information needs of participants are identified. The project emphasizes establishing methods for data integration, from researchers to implementation and documentation, which includes developing a prototype system to support the use of integrated data.

The solution

By involving experts from different partner organizations and adopting standards aimed at cross-domain FAIR implementation, the project aims to enhance visibility into integration methods and processes, effectively addressing multi-disciplinary research questions. Cross-domain data use requires contextualizing metadata, providing a detailed description of data processing from its source form to an integrateable format suitable for researchers.

The provenance description is made accessible to researchers through an integrated browser view, based on the process description capabilities of the DDI - Cross-Domain Integration (DDI-CDI) standard. By providing visibility into data processing at each stage and linking it to descriptions of variables and methods (including the processing code), a better understanding of the resulting integrateable data can be presented. ESS Labs serves as the platform for showcasing these findings.

EOSC Future

EOSC in practice story #22

🏷️ **Keywords:**
#Multidisciplinary, #ClimateNeutral, #SmartCities, #SSHOC, #ENVRIFAIR, #ResearchCommunity, #EOSCinPractice

Useful tips & tricks

Although the project already had an existing help desk system, efforts were made to integrate it with the EOSC Help Desk service. This integration was facilitated through meetings with professional and competent service providers, who explained the available options and assisted in selecting the appropriate one.

Benefits and impact

Society's complex challenges demand collaboration among researchers from various disciplines. The project highlights urban areas as examples, where understanding environmental and physical factors and their interactions with inhabitants' behaviors and attitudes is crucial. Organizational and scientific collaborations across domain boundaries can foster more creative and well-informed policies. The project aims to identify practical ways to enable such collaborations at all levels, from organizational and scientific to technical aspects. By exploring best practices in these areas, the project aims to contribute to the resulting societal benefits.

Why do I need EOSC?

Currently, there is no standard organizational framework for conducting multi-disciplinary research. EOSC offers a structure that facilitates interoperability and collaboration at all levels, becoming an integral part of the work of infrastructure players, data providers, technical implementers, and researchers. Without a framework like EOSC, such collaborations remain ad-hoc and cannot become regular scientific practice.

Across disciplines

The project's inherent focus is on cross-disciplinary science. Through the ESS Labs service, all project findings, including scientific collaboration, methods, data, implementation, metadata models, and programming code, are accessible to serve as an example of how to practically approach multi-disciplinary science.

Useful material related to this story

climate-neutral-and-smart-cities_en

sshopencloud.eu

envri.eu

Future developments

The ESS Labs service envisions adding other prototypes and exemplary projects in the future, involving innovative approaches to social science data and its cross-domain use. Expanding beyond the current environmental and climate data projects, the ESS Labs approach provides a valuable channel for exploring future opportunities in various domains.

Future funding model scenarios

The ESS Labs service serves as a channel to expose further project results, combining reports and documentation with prototype implementations and supporting programming code and data. Future projects must identify their own funding sources and be relevant to the work of the ESS ERIC. Projects of all sizes will be considered based on topic complexity, the number of partners involved, and the degree of implementation, as long as the ESS remains a significant partner and supports the proposed additions to the ESS Labs in line with their current direction.

Creating FAIR Knowledge Graphs for Open Imaging Data Sharing in EOSC/COVID-19 Demonstrator

Leveraging OpenBel, pybel, and EOSC Infrastructures to Facilitate Cross-Domain Research and Reproducible Workflows for Disease Knowledge Graphs

Reagon Karki

Data Manager/Researcher
Fraunhofer ITMP

"The main challenge is to create reproducible workflows for generating Knowledge Graphs (KGs) that enable cross-talk between biological and chemical entities, while also addressing discrepancies in entity names and representations."

The Service Provider

KG Generator

EU-OS/Fraunhofer ITMP takes the responsibility to consolidate and FAIRify data generated at several EU-OS partner sites for Test Science Project titled 'Open Imaging Data Sharing in EOSC/COVID-19 as Demonstrator'. Using these data, Knowledge Graphs embedding chemotype-phenotype of diseases are created.

The Challenge

The main scientific challenge is to develop workflows to enable cross-talk between different biological and chemical entities for creating Knowledge Graphs (KGs). Currently, there are discrepancies in entity names and representations. Moreover, the challenge is also to create reproducible workflows such that KGs can be generated within no time. The technical challenge is to find appropriate infrastructures to create the workflow and deploy it in a suitable platform once the KGs are created.

The user community

The targeted users for our resource are expected to be mostly researchers studying protein activities, chemoinformaticians and data scientists, mostly involved in COVID-19 and infectious diseases. However, in a longer run, the workflow will be applicable for researchers of any disease field.

The solution

The solution to the challenges were achieved by using open-source framework (i.e. OpenBel) and python package (i.e. pybel) for capturing and representing the underlying semantics of the data. Additionally, we have integrated public curated databases such as ChEMBL, UniProt and OpenTargets to expand the knowledge of chemical and proteins on assays, biological process, functions and pathways and adverse effects. The KG is compliant with FAIR annotations allowing seamless transformation and integration to/with other formats and infrastructures. The EGI notebooks from EOSC catalogue of services was used as the workspace and tool to deploy the codes and scripts.

The codes are available in the following github repositories.

1. https://github.com/Fraunhofer-ITMP/BY-COVID-KG
2. https://github.com/Fraunhofer-ITMP/mpox-kg

The links for the resources that are available in EOSC marketplace are as follows:
1. https://explore.eosc-portal.eu/search/software?pid=10.5281%2Fzenodo.7351393
2. https://explore.eosc-portal.eu/search/software?pid=10.5281%2Fzenodo.7351221

EOSC in practice story #23

Keywords:
#Reproducibleworkflows,
#KnowledgeGraphs(KGs), #Biologicalentities,
#Chemicalentities, #Entityrepresentations,

Useful tips & tricks

Although the workflows can be used in mybinder.org, I recommend using the workflows in EGI notebooks. Because, once users sign up for EGI notebook account, they have a personalized workspace to run the scripts. This comes with well supported IT-team in case of issues, easy cloning of github repo, saving workspace anytime and enough space to hold large datasets.

The impact on society

Our main aim of developing the workflows is to make them reproducible such that users can quickly and easily generate KGs for any disease of their interest. One of our main objective is to facilitate researchers with no or little programming skills.

This will enable them to produce KGs with minimal input and make interesting scientific queries to the KG. Moreover, the KGs can be used for downstream analyses such as finding chemicals with similar/ dissimilar structures and identifying active assays for protein of interest.

The deployment of the workflow using EGI notebook will bypass tools installation/set-up, meaning that a cloud-based service can be used for all the activities.

Why do I need EOSC ?

EOSC supports the providers with many different features such as monitoring, accounting and long-term sustainability of the resources. Furthermore, providers can use pre-exisiting tools to deploy their scripts and codes (EGI notebooks in our case). For users, they can go to the marketplace and search any resource with keywords and start using them immediately.

Across disciplines

We are already collaborating and aligning our work with our partners who come from different domains. For example, users of KG can directly access 3D-Bionotes which is a database of protein 3D structures being inside the KG.

Useful material related to this story

COVID-19 Metadata Findability and Interoperability

Future developments

Right now, the use of KG workflow needs to prepare some data by the user. For example, the user needs to have files containing list of proteins or chemicals associated with a disease. We intend to remove this preparation for users and provide a feature to type in name of a disease which will then create a KG instantly.

Sustainability for an EOSC in practice

We would like to see the opportunity to extend the funding period of the project and improve the features of current work.

Future funding model scenarios

The upcoming activities in our current work will be to improve the existing workflow by enabling users to generate KGs at ease with minimal input. Moreover, the deployment of this resource on a cloud-based server will effectively remove the hassle of local installments of tools and packages. We have deployed this in EGI Notebooks: Researchers Edition and are considering upgrading to Community Edition. We have already published a paper highlighting the reproducibility of the workflow using Mpox as a use case. We definitely will publish another paper with the updated and improved features in the KG workflow. Additionally, we aim to offer training and workshops to potential users of our resource.The funding project for our extended work is not yet identified but we see the opportunity to be part of EOSC-Beyond as we seem to align with its scope and objectives.

Dashboard for the State of the Environment

A European Dashboard showcasing the State of the Environment

Angeliki Adamaki

EU Project Manager
ICOS Carbon Portal, Lund University

"The Dashboard's central service is designed and built on a fully open-source foundation, and it embodies a flexible and scalable system, reflecting the core concept. As a result, we encourage other Research Infrastructures and EOSC providers to adopt this service as an additional platform for sharing their data and services with a wider audience."

The Project Involved

ENVRI FAIR

EOSC Future

The Dashboard for the State of the Environment is developed by the Environmental Research Infrastructures (ENVRIs) as one of the Science Projects of the EOSC Future, using the experience and results of the ENVRI-FAIR project. The ENVRI Dashboard aims to make European environmental data accessible to a broad audience through user-oriented indicators displayed in an easy-to-approach interface. The overarching goal of the project is to equip the broader public and policymakers with a user-friendly, customisable dashboard that visualises an extensible collection of indicators that describe the state of the environment in near real-time.

The User Community

ENVRI Community

The Environmental Research Infrastructure (ENVRI) community is a cluster of European research infrastructures focused on the environment and Earth system science. The ENVRI-FAIR project which concluded in June 2023 aimed to advance the FAIRness of their data and services with emphasis on their interoperability and connect the ENVRI community to the European Open Science Cloud (EOSC).

The Challenge

The project set out to find a solution for providing information on the state of the environment displaying multi-disciplinary data from multiple sources in a coherent and user-oriented interface. The developing teams had to create interchangeable and rearrangeable frames, each carrying its own customisation menu. When logged in, user preferences need to be saved for easier and faster access to the preferred data and services, allowing the users to configure their own Dashboard with the indicators of their interest.

The solution

The Science Project demonstrates how different services and multi-disciplinary data resources from multiple research groups can be combined and used to generate scientifically justified environmental indicators in an easy way. The components make use of existing RI services that are onboarded to EOSC, while the new Dashboard service components are integrated in the EOSC for wider access and possible re-use by other research groups. Moreover, several EOSC core services are used, including the EOSC AAI federation, the monitoring service and the EOSC Helpdesk.

The project aims to:
 1. develop and launch a dashboard for environmental indicators by setting up analytical workflows for different environmental disciplines and integrating their outputs;
 2. connect the analytical frameworks to the Dashboard service and integrate them in the EOSC platform;
 3. mobilise and empower larger scientific communities engaging them as providers, co-creators and end-users;
 4. demonstrate and promote the benefits and potential of web-based science using EOSC

EOSC Future

Benefits and impact

The Dashboard for the State of the Environment is a powerful tool that gathers a significant trove of European environmental data in one location. The simplicity of the interface lowers the threshold for non-specialists and citizen scientists to access the data. At the same time, higher-end technical possibilities offer additional options for more advanced users. The Dashboard can serve both scientists and the public by giving access to environmental data and services that are developed at the infrastructure level. The ENVRI services will be developed further within the EOSC environment and the user communities will continue growing. The impact of the project can be seen at strategic, scientific and policy level.

The Dashboard is designed to be completely user configurable so that the users can select from a list the indicators to be shown and their order. Providers can add, remove and edit indicators through a standard REST API, that allows transferring all parameters, including the configuration of the indicators and how to provision data values and thumbnail interaction. The Dashboard is implemented and operated using engineering best practices, including YAML for the indicators' descriptions and a robust and flexible container-based deployment.

Why do I need EOSC?

Through EOSC Future, the Dashboard has been developed using several EOSC services available on the EOSC Platform. As an integration platform that supports scientific workflows, the EOSC provides core services to support research infrastructures to further develop or improve their services, which are also onboarded on the EOSC Platform, and become available to all EOSC users.

The Dashboard is part of the ENVRI-Hub architecture, a common virtual interface for ENVRI-FAIR data and services that can be integrated into higher level platforms as is the EOSC, enabling users from a broader scope to easily access the available data and services.

Useful material related to this story

EOSC Future_Data in action

envri.eu

Across disciplines

To showcase the benefits of an integration platform that supports scientific workflows, the ENVRIs develop the "Dashboard for the State of the Environment" as a cross-RI and cross-discipline service to address with scientific facts the environmental concerns. The project brings together three scientific domains (Atmosphere, Biodiversity, Ocean) that each have set up analytical workflows to provide environmental indicators (in near-real-time wherever possible), allowing the users to visualise the "State of the Environment" by interacting with the service interface. The Dashboard's strength lies in its universal and flexible nature, capable of expanding continuously with its extensible collection of indicators. Moreover, it allows data providers from other clusters and disciplines to contribute additional social/behavioral data, enriching the user's understanding of the environment's state.

Future developments

This Science Project leverages various EOSC services, including AAI (Authentication and Authorization Infrastructure), cloud services, and data storage. The workflows responsible for generating the environmental indicators also integrate with the EOSC and Research Infrastructure (RI) computing capabilities. As a proof of concept, a limited set of indicators is currently accessible. However, it is anticipated that the participating RIs will contribute a plethora of additional indicator options in the coming times. With the appropriate technologies in place (APIs, metadata templates etc.), new RIs will have the opportunity to join the initiative and contribute their own indicators to the "Dashboard for the State of the Environment."

Interdisciplinary Approach to FAIR Metadata: Enabling Collaborative Science in the European Open Science Cloud

Enhancing cross-domain metadata interoperability, integration and impact on COVID-19 research across Science Clusters

Maria Panagiotopoulou

Senior Project Manager at ECRIN

"For the first time, the META-COVID project brought together metadata experts from different research infrastructures in the fields of Life Sciences and Social Sciences and Humanities with the aim of discussing interoperability of information on COVID-19 across domains. The project inventoried the different metadata schemas used by the research infrastructures, discussed the creation of a common "contextual metadata" framework and linked the different metadata catalogues with relevant EOSC work, such as onboarding them in the Marketplace or exploring the applicability of the EOSC Interoperability Framework."

The Project Involved

Within EOSC Future, the Science Project "COVID-19 metadata findability and interoperability in EOSC" (in short META-COVID) aimed to explore the importance of "contextual metadata" in the fields of Life Sciences and Social Sciences and Humanities and to develop a framework for a metadata model characterising the research approach and workflow across research infrastructures. Contextual metadata are referring to a) information about the research process that generated the data, including descriptions of that process and the methodologies used, and b) information about the 'inputs' into the research process – e.g., grants, people, organisations, regulators, and research infrastructures and resources. The project is guided by three practical use cases in the field of COVID-19. One use case looking into enhancing the interoperability of the ECRIN Clinical Research Metadata Repository and the BBMRI-ERIC Directory, another looking into cross-linking information from the CESSDA Data Catalogue and the ECRIN Clinical Research Metadata Repository and a final one focusing on designing an ontology for COVID-19 related topics as discussed in parliamentary debates and social media.

The Research Community

META-COVID involved two Science Clusters: Life Sciences represented by the partners ECRIN, BBMRI-ERIC, EATRIS and EU-Openscreen and Social Sciences and Humanities represented by CESSDA and CLARIN. In order to collect feedback and achieve alignment, META-COVID collaborated with other (EOSC) projects of relevance (e.g. EOSC-Life, BY-COVID, ISIDORe, EOSC4Cancer, FAIRCORE4EOSC, FAIR-IMPACT), members of the EOSC Task Force on Semantic Interoperability, actors involved in the EOSC Interoperability Framework and relevant RDA Working Groups. The outcomes of this project are of interest for researchers working on metadata, ontologies, and vocabularies.

The Challenge

Currently, there are metadata schemas (such as DataCite, or DDI) that can describe the concrete outputs of research, e.g., papers and datasets, but relatively little work has been done on finding a metadata schema for the research itself. The problem is that different disciplines have vastly different ways of organising research activities, for instance because of differences in funding models and mechanisms, or in requirements for approval, and thus differences in how and when research is split into discrete activities and labelled.

In addition, research efforts take place at a variety of scales, have varying requirements for pre-published protocols, use a huge range of different methodologies and workflows, and may even draw upon different underlying assumptions. Research design, approach, strategy and method are heavily influenced by the researchers' epistemology and research philosophy. The type of the research (e.g., hypothesis testing versus hypothesis generating), the methodology chosen (e.g., experimental, survey, cohort, case study) and the research methods applied (e.g., type of sampling) are of major importance in understanding the data generated, and thus in supporting any secondary use of that data. Another issue to be solved is the integration of various sources of information related to parliamentary and social media metadata. Consequently, metadata should go beyond a description of the data itself to include the basic elements of the research process ("contextual metadata").

The solution

The COVID-19 pandemic has generated a huge variety of research activities, studies and policies across both life sciences and social sciences and humanities: examples include genomic sequencing, assays of immune response, clinical trials, population health analyses, exploring vaccine hesitancy, investigating the role of social media, public debate and economic analyses of the impact of public policy issues (e.g., lockdown measures, imposed face masking). Potential insights from combining the data and conclusions from these different forms of research are, however, made more difficult by the lack of a common metadata framework with which to describe them.

EOSC Future

EOSC in practice story #27

Keywords:
#EOSC, #FAIR, #Interdisciplinary,
#KnowledgeGraphs, #COVID19, #COVID19data,
#ResearchCommunity, #EOSCinPractice

Even within one scientific cluster (e.g., SSHOC, EOSC-Life), the metadata landscape is heterogeneous and numerous domain-specific standards are applied (e.g. MIABIS for biobank information, DICOM for images). Developing widely applicable metadata is a key part of rendering data more valuable, by allowing them to be more easily found and characterised, regardless of the discipline in which they were generated. META-COVID inventoried the different metadata schemas used by the research infrastructures and generated a proposal for a framework to describe contextual metadata. As part of the work, Knowledge Graphs for COVID-19 and for Monkeypox have been developed to represent complex scientific information.

EOSC service or tool used

The project leveraged the EOSC as a unique environment to collaborate with metadata experts from different scientific domains. The work around the EOSC Interoperability Framework impacted META-COVID and got impacted by it. In addition, a lot of the partners' metadata catalogues and tools have been made available through the EOSC marketplace, ensuring easy accessibility for users. Service providers benefited from receiving information about total views and downloads, enabling them to monitor the impact of their resources effectively. Project partners became also aware of and embedded in their work other products from the EOSC Marketplace such as the EGI-Notebooks, a horizontal service within EOSC, which allowed developers to work on personalized and secure remote servers without the need to install any software/tools on their local PCs.

Benefits and impact

META-COVID allowed for the first time a multidisciplinary discussion among metadata experts from different infrastructures in the Life Sciences and Social Sciences and Humanities on the FAIRness of metadata in their respective fields and interoperability issues across domains. The use cases within META-COVID will improve the discoverability of data objects related to COVID-19 across different research infrastructures. For example, ECRIN and BBMRI-ERIC are enriching their metadata catalogues so that information on biosamples can be linked to information about the clinical study that generated them.

Useful material related to this story

ECRIN

The development of a COVID-19 ontology to describe information from parliamentary debates and social media will also strengthen our knowledge on the impact of the different health policy measures on the population and lead to the informed design of future biomedical studies. The contextual metadata framework will inform other relevant initiatives such as the European COVID-19 data portal.

Why do I need EOSC?

EOSC is creating a multidisciplinary environment of Open Science practices, tools, data and services in which metadata interoperability across different scientific fields plays a crucial role for the findability of relevant data objects. EOSC facilitated collaboration and provided a platform for sharing resources and knowledge. It allowed the project team to host their resources in a centralised marketplace, making them easily accessible to the research community. Furthermore, EOSC services like EGI-Notebooks eliminated the need for local installations and provided a secure and personalised remote server environment for efficient and cost-effective knowledge graph creation.

Across disciplines

META-COVID is interdisciplinary by design. The project's approach of capturing "contextual" metadata was finetuned with diverse research infrastructures, spanning clinical research, biobanks, chemical biology, social media and parliamentary talks, social science surveys. Although the framework has been developed with COVID-19 in mind, there is nothing impeding its generalizability to other disease areas.

Innovating HVAC Industry with AI Models: The PreMaCOOL Journey

Leveraging EOSC Services for Energy Forecasting and Maintenance Solutions

Thanasis Anagnostis

Research, Development and Innovation Manager at Klimamichaniki
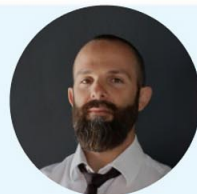
"The development of AI models for energy forecasting and predictive maintenance is innovative for the specific domain, enabling Klimamichaniki to gain new knowledge during the EOSC DIH pilot and eventually develop innovative complementary services for industrial installations."

The Project Involved

PreMa COOL

Klimamichaniki was founded in Thessaloniki, Greece in 1984. It is an engineering consultancy and construction company with expertise on all HVAC systems. The company recently pivoted towards research and development via incorporating and applying AI methods within its modus operandi, to develop innovative solutions.
https://eosc-dih.eu/premacool/

The Challenge

For interoperable data services, explicitly-defined data variables are crucial for usability, aligning with FAIR principles. Poorly-aligned semantic resources, though, complicate cross-disciplinary use. Despite progress, the terminology field faces new challenges like multilingualism, sustainable concepts, and reuse. Unified strategies are needed for vocabulary definition across research areas. Challenges include completeness variations, monolithic frameworks, and limited curation. Software designs for semantic artifacts must tackle data quality requirements

The Research Community

We are a commercial company in engineering and constructing, recently we are also into research and development via incorporating and applying AI methods within its modus operandi, to develop innovative solutions.

EOSC service or tool used

- OpenAIRE EXPLORE: Academic research - OpenAIRE NEXUS https://explore.openaire.eu/
- EGI Notebook: Develop code - EGI Foundation (EGI-ACE) https://www.egi.eu/service/notebooks/
- DEEP training facility: Train models - https://deep-hybrid-datacloud.eu/the-platform/
- ARGOS: Create data management plan - OpenAIRE NEXUS https://argos.openaire.eu/splash
- B2SHARE: Networking and dissemination of results - EUDat (DICE Project) https://www.eudat.eu/catalogue/b2share

PreMaCOOL has a sequential implementation plan, common in most research works, study the bibliography and market, try the solution in mind, adhere to any obligations and restrictions and finally share the results. Therefore the logical way to implement EOSC services was to utilize OpenAIRE EXPLORE for academic research, and use search engines to search for similar existing solutions. Then, access to tools and resources were required to develop code, implement ideas, train models, and test their results., therefore EGI Notebook and Deep Training facility were a perfect fit. In fact, most of the PreMaCOOL effort was spent using these two services. Then, PreMaCOOL utilized ARGOS to create a data management plan, which is necessary since the data that were used were not public therefore a data management plan might come useful in the future. Finally, B2SHARE was selected to disseminate the results of PreMaCOOL. This has not been used yet, however, it is within the plan.

EOSC Future

Useful tips & tricks

PreMaCOOL has a sequential implementation plan, common in most research works, study the bibliography and market, try the solution in mind, adhere to any obligations and restrictions and finally share the results. Therefore the logical way to implement EOSC services was to utilize OpenAIRE EXPLORE for academic research, and use search engines to search for similar existing solutions. Then, access to tools and resources were required to develop code, implement ideas, train models, and test their results., therefore EGI Notebook and Deep Training facility were a perfect fit. In fact, most of the PreMaCOOL effort was spent using these two services. Then, PreMaCOOL utilized ARGOS to create a data management plan, which is necessary since the data that were used were not public therefore a data management plan might come useful in the future. Finally, B2SHARE was selected to disseminate the results of PreMaCOOL. This has not been used yet, however, it is within the plan.

Benefits and impact

Within PreMaCOOL, Klimamichaniki has accomplished a set of achievements:• Developed a set of anomaly detection models based on point-wise (One-class SVM, Local Outlier Factor, Isolation Forest, Autoencoders) and patter-wise (Hidden Markov Models, LSTM autoencoders, CNN autoencoders) detection mechanism. Visual inspection on the data showed accurate detection of present anomalies.• Developed a set of energy forecasting models with hourly granularity and forecasting depth of 24 hours, aiming to predictict next day's consumption patterns. The selected algorithms (SARIMA, LSTM, Transformers, Neural Prophet) managed to achieve mean square errors below 0.01.• Through the engagement with the pilot and EOSC DIH, two new ideas were conceptualized on innovative solutions based on data and AI methodologies. One of them has been submitted for funding via the funding calls proposed by EOSC DIH.• The competence of HVAC engineers, towards identifying potential data-driven solutions to existing market problems, has been significantly increased. The company's engineers are now is a continuous engagement and discussions on whether implementation of data-driven AI solutions could potentially address a number of specific problems in the domain.

Why do I need EOSC?

EOSC provided the services and infrastructure that was crucial for Klimamichaniki to proceed with the creating of a strategy and the training AI models at this stage of development. Since the company is not a software company and has only recently started being involved in innovation concepts, EOSC provided with the tools and the human support to guide us through this endeavour.

Across disciplines

Development of machine learning models is not an easy task for non-experts. Klimamichaniki has vast experience in the energy and HVAC domain, but is new to implementing machine learning and artificial intelligence algorithms. We hope our experiences will be valuable for the R&D departments of other SME's and start-ups who want to start with AI and ML.

Limitations and future improvement

Given that PreMaCOOL has reached a satisfactory point and some promising results have been produced, two services that could be useful at this stage (but I'm not sure if they were available at the time of submission), are:production-level software and operations: a set of tools and experts that could enable the SMEs to bring their early-stage solution and make it a production-level product with software engineering techniques and DevOps. Especially for AI and ML project, which is very common lately, MLOps and deployment would be also included.business consultancy: a set of experts that would create potential business ideas from the developed product, and help the SME to select the best one based on the potential and the market. This would also include exploitation plans and possibly scale-up possibilities.

Useful material related to this story

eosc-dih.eu/premacool/

Searching for bright compact needles in a radio haystack of galaxies and stars

Unlocking the Mysteries of Fast Radio Bursts (#FRBs) and Compact Radio Sources through Innovative Data Analysis and EOSC Integration

Dany Vohl

Research Associate, University of Amsterdam, Affiliated scientific researcher; ASTRON, Netherlands Institute for Radio Astronomy

"The FRB community has been steadily growing since 2007, having a major annual conference, along with regular meetings and workshops. The field now covers various aspects ranging from theoretical modelling of likely progenitors and their emission mechanisms; observational methods, radio telescopes design, and algorithmic advancements; studies of host environments to FRBs; multi-wavelength and multi-messenger studies; and using FRBs as probes of the interstellar medium, circum-galactic medium, and potentially to cosmology."

The Project Involved

The discovery of fast radio bursts (FRBs) is one of the most intriguing radio astronomical discoveries of our time. FRBs are short (faster than the blink of an eye) and energetic (as much energy in a few milliseconds as the Sun outputs in a day) radio flashes originating from outside our Milky Way galaxy. The nature of FRBs remains elusive, but causality arguments based on the short duration of these bursts imply a compact origin. Furthermore, a detection of FRB-like emission from the Galactic magnetar SGR 1935+2154 hints towards a neutron star origin. From the handful of FRBs that have been localized to a host galaxy, two published cases have been associated with compact, persistent radio sources (PRSs). Both display a characteristic broadband synchrotron spectrum and both dwarf galaxies (containing less than about three billion solar masses), sharing similarities with intermediate-mass black hole (IMBH, 100 to 100,000 solar masses) candidates. To further our understanding of the role of PRSs in the lifecycle of FRBs, and potentially black hole growth, it is crucial to discover more of them.

The Solution

To search for PRSs and IMBH, our team developed the matchmaker framework. Matchmaker is designed as an intuitive way to cross-investigate information from various astronomical catalogues. The framework is open-source and written in the Python language, making it simple to extend and use along other community-based packages. In particular, we make use of available Virtual Observatory packages where possible. The use of Python also makes matchmaker easy to port to various computing facilities, like those provided on EOSC Marketplace (e.g. Notebook servers). With minimal effort, one can start matching information between catalogues and rapidly investigate scientific questions. For our purpose, we searched the millions of radio sources contained in the second data release LOFAR Two-Meter Sky Survey (LoTSS) against other catalogues at different wavelengths (mid-infrared, optical, ultraviolet, X-ray, and gamma rays) and identified about 30 PRS/IMBH candidates.

The Research Community

The FRB community has been steadily growing since 2007, having a major annual conference, along with regular meetings and workshops. The field now covers various aspects ranging from theoretical modelling of likely progenitors and their emission mechanisms; observational methods, radio telescopes design, and algorithmic advancements; studies of host environments to FRBs; multi-wavelength and multi-messenger studies; and using FRBs as probes of the interstellar medium, circum-galactic medium, and potentially to cosmology.

Why do I need EOSC?

EOSC provides an ecosystem of data, software and publications. For instance, EOSC marketplace simplifies the entry-point for both researchers and students wanting to perform analyses on computing facilities. Within a few clicks, researchers can gain access to top-of-the-line compute and data, investigate research questions, or even reproduce scientific results. EOSC acts as a catalyst, enabling researchers at various career levels (from students to professors) to perform decentralized research.

EOSC Future

EOSC service or tool used

To achieve its goals, the project leveraged the capabilities of the European Open Science Cloud (EOSC), particularly utilizing the EOSC marketplace as a centralized platform.

This platform serves as a one-stop-shop that provides the means to discover and share a wide range of resources, including software, services, data sets, and publications.

Through the marketplace, we could access computing services, initiating data exploration and analysis. Logging in with AAI credentials provides access to a wealth of computing resources and datasets, making it possible to analyze and process data efficiently.

One of the key EOSC services used in the project was the EDI Notebooks service. From within marketplace, upon selecting the EDI Notebooks service, we could request the appropriate bundle for a given task, such as the "EDI Notebooks + B2Drop", which provided access to both computing resources and storage capabilities.

The seamless integration of these services within the EOSC framework allowed us to streamline the workflow and focus on scientific endeavors.

Benefits and impact

The benefits and impact of the project were various. Firstly, FAIRness of astronomical data and EOSC services empowered us to gain a deeper understanding of the potential population of luminous compact radio sources.

The synergy between data access, compute resources and human resources across EOSC facilitated comprehensive analyses, leading to insights that were not easily attainable otherwise, and sharing of method and results for better reproducibility.

Furthermore, the EOSC Portal and Marketplace provided a place to showcase our research and framework to a wider community and to engage with prospective users.

Useful material related to this story

matchmaker

Across disciplines

EOSC marketplace allows researchers from various scientific background to find datasets from other fields. While our software framework has been developed in the context of Astronomy and Astrophysics, the mechanisms it provides to deal with heterogenous datasets can serve researchers from other fields as well after minimal adjustment tailored for their science case.

Furthermore, cloud computing resources offered through EOSC can allow cross-disciplinary groups to form by accessing shared resources and storage. These research activity methodologies can serve as a platform to ease fluid collaborations.

Bringing the global research data community and data initiatives closer to EOSC

Stories from the RDA/EOSC Future Open Calls

Collaborative Knowledge Sharing: a case study on EU FAR project

Leveraging Zenodo, RoHub, RDA, and EOSC Marketplace for data sharing and collaboration

Monica Marin

Researcher at the Research Institute for Quality of Life, Romanian Academy, EU FAR

"EOSC services and tools provide a wider dissemination for project outputs. The key idea of open science is therefore promoted in a research ecosystem where re-sharing of data is not always the rule. Larger dissemination can provide the impetus to build similar databases in other EU countries."

The EOSCFuture.eu/RDA open calls

The open calls managed by the Research Data Alliance (RDA) with funding from EOSCFuture.eu aim to bring data initiatives closer to EOSC and provide EOSC with expertise, tools, best practices, and standards from the global research data community, with a total million Euro reserved to fund over 65 grantees over 12 open calls.

The Project Involved

EU FAR

EU FAR – EU Funds by Area Results - was supported under the RDA Open Call for Cross Disciplinary Science Adoption Grants #1 and it was implemented by the Research Institute for Quality of Life from September 2022 until April 2023. The key objective of the project was to stimulate the development of open data on EU funding, in line with FAIR principles (Findability, Accessibility, Interoperability and Reusability). Information publicly displayed by the project is focused on EU funds absorbed by local governments in Romania (LAU 2 level) from 2014 to 2020, with a special focus on rural areas.

The Research Community

The EU FAR user community is mainly composed of scholars from sociology, public management, geography and economics as well as academics interested in general in open science practices. EU FAR reuses open governmental data; consequently, stakeholders from public sphere will be interested in the increased re-usage of data from public institutions.

The Challenge

The main challenge EU FAR addresses is the lack of granular data on EU funding. In Romania, open governmental data still include errors on key aggregators (mainly LAU 2 code) and are not fully aligned with the FAIR movement. Additionally, a thorough methodological approach is needed to aggregate data between budgetary years and have a comprehensive overall picture of what has actually happened with EU funding at the lowest administrative level, that of the municipality. Part of these challenges are valid not only for Romania, but also for other EU countries. Especially in post-communist countries, territorial disparities remain a significant issue, with rural areas placed at a disadvantage on multiple dimensions.

EOSC service or tool used

All project deliverables are available on Zenodo, and one of the key deliverables of the project – EU FAR database – is available on RoHub as well in ČSDA – Czech Social Data Archive, which is part of CESSDA. Additionally, project outputs are also available on Open Science Framework and Scipedia, which are also promoted in various research available from EOSC Marketplace. We found very useful ideas in training materials available from EOSC, on various topics: open science, research data, FAIR sharing, and FAIR promotion.

Useful tips & tricks

We initially searched EOSC services and tools when submitting the research proposal. However, in the meantime, EOSC had evolved, and we were happy to discover better search options as well as the availability of useful training materials. Therefore, being constantly informed about EOSC pays off. Additionally, we would like to express our sincere thanks to the Czech Social Data Archive for agreeing to publish our database, as the Romanian partner from CESSDA is no longer functional. Being now part of CESSDA greatly increases the discoverability of the EU FAR database.

Benefits and impact

EOSC services and tools provide a wider dissemination for project outputs. The key idea of open science is therefore promoted in a research ecosystem where re-sharing of data is not always the rule. Larger dissemination can provide the impetus to build similar databases in other EU countries. The EU FAR database aggregates public open data on the volume of EU funds absorbed by each municipality (LAU 2) level in Romania for the entire programming period of 2014-2020. Data are converted into euro at constant 2010 prices per inhabitant to provide comparable data both between Romanian municipalities and, eventually, other European territorial administrative units. Maps at national, regional (NUTS 2) or county (NUTS 3) levels illustrate the results on EU funding. Analysis of spatial disparities between rural areas is part of a working paper. Additionally, a policy brief in plain language explains which are the key areas at stake and the needed policy actions.

Why do I need RDA?

We would also like to mention the role of RDA: the plenary meeting and its co-located events provided ample opportunities for networking, brainstorming and exchanging lessons learned. The RDA Overview for the Social Sciences Report (October 2022) and the Metadata Standards Catalog WG Final Report were also very useful documents in the planning stages of the project.

Useful material related to this story

eoscfuture-grants.eu

rda-alliance.org

Why do I need EOSC?

EOSC represents a platform for the wide dissemination of research results. It is highly useful in connecting researchers from various disciplines and countries, which is key to further developing the EU FAR database with similar initiatives across EU countries. By showcasing different research results from our projects or other relevant projects, EOSC services and tools promote open access to research products and reduce the risk of duplication of efforts, which is critical to advancing the research agenda.

Across disciplines

EU FAR is a cross-disciplinary project, with research results that are valuable for sociology, economics, geography and public management. The project team aims to further facilitate cross-fertilization of results by seeking to also further improve the reusability part of open governmental data.

Limitations and future improvement

As publications represent one of the key indicators for evaluating researchers, better cross-linkages between data repositories and publications repositories would be very valuable. Some of the preprint and journal platforms have already done this, but further improvements would be beneficial for the research community in general.

EOSC Future

RDA RESEARCH DATA ALLIANCE

EOSC Future/RDA Open Calls

EOSC in practice story #24

Keywords:
#data, #OpenScience, #FAIRprinciples, #FAIRsharing, #FAIRdata #Trust, #FAIR indicators, #EOSCinPractice

Enhancing FAIRness and TRUSTworthiness in the REPOPSI open repository

Leveraging EOSC and RDA outputs to improve the Open Repository for Psychological Research Instruments (REPOPSI)

Aleksandra Lazić

Research Assistant at the University of Belgrade – Faculty of Philosophy, Serbia

"The EOSC Portal and Marketplace have enabled us to showcase our work to a broader user base and establish connections with potential users in new ways. Additionally, EOSC has elevated our visibility and credibility at a European level, with the potential for REPOPSI to make an impact on the global community"

The EOSCFuture.eu/RDA open calls

The open calls managed by the Research Data Alliance (RDA) with funding from EOSCFuture.eu aim to bring data initiatives closer to EOSC and provide EOSC with expertise, tools, best practices, and standards from the global research data community, with a total million Euro reserved to fund over 65 grantees over 12 open calls.

The Project Involved

REPOPSI was supported under the RDA Open Call for Cross Disciplinary Science Adoption Grants #1 and it was implemented by the Laboratory for Research of Individual Differences at the Faculty of Philosophy of University of Belgrade in Serbia from July 2022 until July 2023.
The key objective of the project was to improve the TRUSTworthiness and FAIRness of the REPOPSI (meta)data by implementing RDA recommendations and other outputs . The project focused on building the technologies needed by the repository to improve its adherence to the principles of User Focus, Sustainability, Findability, and Interoperability. Dissemination and community activities were also undertaken to promote further uptake of RDA outputs.

The Research Community

The REPOPSI community are researchers and students in psychology, but also in other social and behavioural sciences. Moreover, REPOPSI is used in teaching university courses in psychometrics and related fields. As currently half of instrument records are available in both Serbian and English, REPOPSI is useful to both local and international researchers

The Challenge

As the first centralised repository for psychological instruments in Serbian, one of the main goals of the project was to prevent duplication of efforts in translating/adapting instruments and to provide a space for local researchers to promote their original instruments. A further challenge was to facilitate easy access to essential materials for researchers and students. While the Open Science Framework - the free platform hosting REPOPSI - is a valuable hosting platform, this also posed challenges in terms of Reusability (e.g. metadata cannot refer to a machine-understandable reuse licence) and Interoperability (e.g. metadata values cannot be expressed using a controlled vocabulary). REPOSI also had the need to promote its data and services to a wider, global research community.

EOSC service or tool used

The team used the "FAIR Data Maturity Model" and "CoreTrustSeal Requirements 2023-2025" to self-assess FAIRness and TRUSTworthiness at the beginning and at the end of the project. FAIRsharing was used to find a FAIR-compliant controlled vocabulary and a metadata schema that could be applied ; from which the Library of Congress Subject Headings and the DataCite Metadata Schema were chosen. Further, REPOPSI was added as a new database record in FAIRsharing and project deliverables are available on Zenodo. REPOPSI is listed in the re3data registry. After registering the institution as a new EOSC Provider, REPOPSI was also added to the EOSC providers hub and can be visited in the EOSC Marketplace here.

Useful tips & tricks

We consulted the following online course while registering our institution as a new EOSC Provider and onboarding REPOPSI as a new EOSC Service: "For EOSC Providers - Onboarding". In the future, we plan to explore other training resources provided by the EOSC Knowledge Hub (e.g. on how to integrate our service with the EOSC Helpdesk or the EOSC Monitoring Service).

Benefits and impact

The EOSC Portal and Marketplace enable us to showcase REPOPSI to a wider user base and to engage with prospective users in new ways. For example, users can now easily compare REPOPSI to similar services and include it in their EOSC Marketplace Projects, along with other resources of interest. EOSC helps researchers perform better, more efficient and transparent science by making it easier for them to find, share, and reuse open data. In addition, EOSC nurtures a diverse, equal, and inclusive research environment by allowing anyone to browse, discover or contribute to the tools and services. Moreover, researchers and innovators who come from developing countries or regions that have been underrepresented in science are able to demonstrate that their services meet EOSC quality standards, expanding their impact in the global community.

Why do I need RDA?

The project aimed to improve adherence to the principles of User Focus, Sustainability, Findability, and Interoperability and the RDA outputs used were ket to its success. Dissemination and community activities were also undertaken to promote further uptake of RDA outputs.

Why do I need EOSC?

The EOSC Marketplace enabled us to browse the research publications and services needed to improve the FAIRness and TRUSTworthiness of our Repository, especially in terms of Reusability and Interoperability, with ease. Ultimately, this allowed us to promote REPOPSI to a more diverse pool of potential users – we could onboard it into EOSC and include it in registered repositories (e.g. FAIRsharing). As the REPOPSI project deliverables are deposited in Zenodo, they can all also be accessed from EOSC. Thanks to its presence in the EOSC Marketplace, REPOPSI can gain more visibility and credibility.

Across disciplines

REPOPSI's structure and documentation as well as the REPOPSI project deliverables are all public and transparent and available in English. Therefore, REPOPSI can serve as a role model for development of other smaller-scale open repositories of existing research data. Furthermore, the methodology developed for REPOPSI can be extended to other scientific domains.

Useful material related to this story

eoscfuture-grants.eu

rda-alliance.org

EOSC Future

EOSC Future/RDA Open Calls

RDA
RESEARCH DATA ALLIANCE

EOSC in practice story #26

Keywords:
#Persistent Identifier, #Interoperability, #Metadata, #FAIR, #RDM #OpenScience, #EOSCinPractice #RDA

In Vitro Veritas: data and metadata interoperability for physical samples

Leveraging the new IGSN (International Generic Sample Number) ID provided by DataCite in a research tool.

Rory Macneil
CEO of Research Space, UK, host of the FAIR Data Podcast

Xiaoli Chen
Project Lead at DataCite

"The funding provided gave the opportunity to enhance interoperability in research data management, guided by the implementation of persistent identification and metadata features utilizing IGSN DOIs from DataCite and the RSpace digital research platform to create seamless data exchange and support streamlined sample management workflows"

The EOSCFuture.eu/RDA open calls

The open calls managed by the Research Data Alliance (RDA) with funding from EOSCFuture.eu aim to bring data initiatives closer to EOSC and provide EOSC with expertise, tools, best practices, and standards from the global research data community, with a total million Euro reserved to fund over 65 grantees over 12 open calls.

The Project Involved

The Project was supported under the RDA Open Call for Interoperability Framework Contributions and implemented by Research Space (RSpace) and DataCite from December 2022 until July 2023. RSpace provided the digital research platform, which features an integrated electronic lab notebook and sample management system that are connected to an ecosystem of tools & services that researchers and data librarians commonly use. DataCite connects research outputs and resources with metadata—from samples and images to data and preprints - and facilitates the creation and management of persistent identifiers (PIDs), integrates services to improve research workflows, and facilitates the discovery and reuse of research outputs and resources.

The key objective of the project was to provide guiding principles for implementing persistent identification and metadata features on research tools to boost interoperability of research data and to support sample management workflows.

The project focused on building interoperability in research tools using IGSN DOIS provided by DataCite, and the RSpace digital research platform, as a case study.

The project work was closely related to the work being done by the RDA Working with PIDs in Tools IG. 50% of the IG co-chairs were involved in the project. A presentation of the project will be made at the joint Working with PIDS in Tools IG and National PIDS Strategies WG session at Plenary21, and the Working with PIDs in Tools IG will consider adopting the project report as an IG output.

The Research Community

Multiple scientific research domains which use tools to create, collect, manage, analyze, store and share research data and documents were involved. E.g., biology, chemistry, ecology, geology, environmental science, and engineering.

The Challenge

Lack of interoperability between tools/e-infrastructures hinders streamlining processes throughout the research lifecycle. It hampers research data and metadata collection and limits the scope for passing this data and metadata on to and from data repositories.

This issue is particularly acute when dealing with data and metadata for physical samples, which presents challenges at two levels. First, creating, collecting, and managing data related to the samples requires carefully designed software with a rich feature set which is flexible enough to address multiple kinds of workflows, yet intuitive for users. Second, sample management tools currently lack support for associating sample data with both ontologies and PIDs metadata.

EOSC Future/RDA Open Calls

EOSC in practice story #26

RDA — RESEARCH DATA ALLIANCE

🏷 **Keywords:**
#data, #OpenScience, #FAIRprinciples,
#FAIRsharing, #FAIRdata #Trust, #FAIR indicators,
#EOSCinPractice

Useful tips & tricks

The following tips are for research providers who are looking to integrate PIDs workflows into their platforms.

Select appropriate research institutions and researchers to engage with. Gather feedback and requirements from institutions and researchers as part of an iterative process. Consult widely with people with relevant and expert knowledge of the infrastructures and workflows being addressed, inside and outside your organization(s). When translating requirements into product design, start with a long list and refine this into a practical, doable minimum proof of concept to start with.

Benefits and impact

The ability to associate IGSNS with sample data has a potentially revolutionary impact on the discoverability of sample data, bringing a range of benefits, including wider dissemination of sample data, conducting more effective experiments, reproducibility of research, and knowledge generation. This benefit extends to the individual researchers and groups who conduct research involving physical samples, and also the institutions they work for. It will also extend to researchers and institutions who were not involved in the work but who will be able to more easily discover information about samples and the experiments they were used in, which will be useful in their own research.

Why do I need RDA?

Representatives of both organisations involved in this project are very active in the RDA, serving as co-chairs and members of many working groups and interest groups, attending Plenaries and presenting at sessions, and in other ways. The group activities and the plenaries provide a critical avenue for learning about important trends in research data management and meeting and engaging with stakeholders of the community, and shape policy and practice. No other research data organization has the breadth of engagement and participation, the depth of knowledge, or the sustained and regular activity of the RDA which is a unique and invaluable resource.

Useful material related to this story

eoscfuture-grants.eu

rda-alliance.org

Why do I need EOSC?

EOSC funding was instrumental in making this research possible, and the feedback received in the early stage of the process was useful, e.g., we decided to focus only on IGSNS rather than PIDS more generally, which would not have been achievable in the short time available. The encouragement we received and the opportunity to present and discuss progress was also very helpful.

Across disciplines

Association of IGSNS with physical samples collected in multiple domains enables wider discovery of sample data between domains, resulting in more opportunities for cross-disciplinary data comparison and experimental exploration.

EOSC service or tool used

The EOSCFuture.eu/RDA grant provided the opportunity for DataCite and RSpace to team up to address the challenge of data and metadata interoperability for physical samples. It obliged the team to set up a specific work programme, and to work to a specified timeline. Further, regular EOSC-RDA group meetings were held during the project giving us insights and guidance on how to take this forward and how to narrow our focus to make the objectives we set achievable.

EOSC Future

RDA RESEARCH DATA ALLIANCE

Guiding Infrastructure Governance And Controlled Vocabularies Requirement (GIGAR-V)

A concerted effort by the TU Wien Library (AT), ERINHA AISBL (FR), DataCove (AT), Natural History Museum Vienna (AT), reNEW Copenhagen University (DK).

Chris Schubert

Head of Media Management and Library-IT, Vienna University of Technology - Library

"Through collaborative efforts within our dedicated team and in alignment with EOSC, we've endeavored to bridge cross-disciplinary gaps and establish a standardized semantic framework. Our work acknowledges the diverse needs of experts from various fields, aiming to provide a transparent and unified solution for the evolving landscape of data analysis and AI advancements."

The EOSCFuture.eu/RDA open calls

The open calls managed by the Research Data Alliance (RDA) with funding from EOSCFuture.eu aim to bring data initiatives closer to EOSC and provide EOSC with expertise, tools, best practices, and standards from the global research data community, with a total million Euro reserved to fund over 65 grantees over 12 open calls.

The Project Involved

The Project was supported under the second RDA Open Call for cross disciplinary science adoption grants and implemented by a consortia comprising the TU Wien Library (AT), ERINHA AISBL (FR), DataCove (AT), Natural History Museum Vienna (AT), reNEW Copenhagen University (DK).

The GIGAR-V team involved experts in terminologies, vocabulary management, registry software, and APIs, as well as members of standardization organizations with a deep passion for interoperable data and the data service landscape. In addition to their collaborative efforts on semantic artifacts and quality management, they collaborated across various scientific domains, including Biodiversity, Spatial Data (Environmental & Earth Observation), Informatics, Stem Cell & Pathogenic Medicine, and Language Models for Text Mining.

The key goal of the project revolved around tackling the intricacies of incorporating "individual" extensions into existing vocabularies while maintaining versioning, provenance, and quality assessment. The project's software functionalities were carefully designed to seamlessly facilitate this holistic approach. Comparison and collection of requirements will validate this and yield valuable generic recommendations for the scientific community.

The Research Community

As the GIGAR-V team members come from diverse and unique scientific disciplines, they may have different backgrounds and areas of expertise. Despite these differences, they have similar needs and encounter analogous challenges are when dealing with semantic artifacts (such as terminologies, vocabularies, etc.). The use of these semantic components fosters a shared interest in terms of learning from and sharing knowledge and best practices across respective disciplines. However, the target audience pertains to three main groups: a) data creators, who receive guidance on enhancing the interoperability of their data by employing controlled vocabulary and published terminologies; b) scientific domain ambassadors, for example, standardization organizations or efforts; c) software developers, who have incorporated technical specifications and requirements to ensure alignment with domain necessities and the established standards from organizations like W3C and technical committees such as ISO.

The Challenge

For interoperable data services, explicitly-defined data variables are crucial for usability, aligning with FAIR principles. Poorly-aligned semantic resources, though, complicate cross-disciplinary use. Despite progress, the terminology field faces new challenges like multilingualism, sustainable concepts, and reuse. Unified strategies are needed for vocabulary definition across research areas. Challenges include completeness variations, monolithic frameworks, and limited curation. Software designs for semantic artifacts must tackle data quality requirements

EOSC Future

RDA RESEARCH DATA ALLIANCE

Useful tips & tricks

The main aim of this project is to enhance recommendations for a software registry concept used in scripts and implementations, while considering cross-domain requirements. Solutions deployed aim to:

Streamline reference concept assignment, including vocabulary registration, curation, and provenance;
Facilitate terminology co-creation with a clear governance framework;
Enable W3C-compliant knowledge representation in ontologies;
Enhance visibility and accessibility of interconnected research data through Knowledge Graphs

Benefits and impact

EOSC establishes a robust research data infrastructure environment through its dedicated research and implementation agenda. From our standpoint, the technical implementation of software and management practices yields a significant advantage: the tools and services developed within EOSC are not confined to R&I projects; instead, they find broader utility in cross-disciplinary contexts and align with various infrastructure services. This extends to EU Data Spaces, Global Earth Observation Monitoring, and the reuse of small-scale, in situ data. The process of garnering consensus on technical specifications and requirements involves effective communication and collaborative discussions with research initiatives like RDA and CODATA. This approach, in our view, is commendably transparent and open.

Why do I need RDA?

As an independent, global platform RDA provides a forum and inclusive mindset to share experiences, discover and provide knowledge, and show case a broad range of research data and infrastructure. RDA also covers aspects of the existing standards landscape and detailed issue statements of daily scientific tasks. Personally, I've received several supporting actions from RDA, like the RDA Adoption Stories, Bloggs, Webinars, as well as the Plenary stage, which helps my team get involved and take these benefits back to our "small" community in Austria.

Useful material related to this story

eoscfuture-grants.eu

rda-alliance.org

Why do I need EOSC?

EOSC establishes a resilient and familiar research data infrastructure environment through its dedicated research and implementation agenda. From my perspective, the advantageous outcome of the technical software implementation and management rules within EOSC is that the focus of tool and service development extends beyond R&I projects, encompassing a significantly broader cross-disciplinary usage and consensus that extends to other infrastructure services like EU Data Spaces. Achieving agreement on technical specifications and requirements involves effective communication and collaborative discussions with research initiatives such as RDA and CODATA. In my view, this approach is commendably transparent and open.

Across disciplines

The essence of the GIGAR-V initiative lies in acknowledging cross-disciplinary factors. Our focus revolves around terminologies, which, within this context, entails establishing semantic content that inherently remains confined to specific domains. Professionals such as Earth scientists, biologists, and chemists create and maintain their semantic resources based on individual preferences. GIGAR-V aims to comprehend these unique requirements, pinpoint shared needs and challenges, and thereby encompass existing standards concerning registry items.
As the utilization of cross-disciplinary data grows – by data analysts, language models, or through techniques like text mining, machine learning, and the burgeoning AI field – the demand for a consistent and transparent semantic layer becomes imperative. This is precisely where terminologies step in to fulfill this role..

Enhancing COVID-19 Metadata Findability and Interoperability

Empowering Research and Policy Through Enhanced COVID-19 Metadata Interoperability

John Shepherdson

Cloud Platform Delivery Director at CESSDA ERIC

"A cross-domain ontology has been developed for COVID-related topics as discussed in parliamentary data, social media data, and integrated into EOSC as a resource that can make data comparable, interpretable and highly communicative for researchers, as well as journalists, NGOs and citizens."

The Projects Involved

SSH Open Cloud (https://sshopencloud.eu/)
Revolutionizing social sciences and humanities (SSH) research, SSH Open Cloud seamlessly merges cutting-edge tech with SSH challenges. Empowering scholars and practitioners, it fosters collaboration, data sharing, and innovative analysis across disciplines. Its robust infrastructure enables users to unlock data potential, fueling insights and informed decisions. Guided by open science principles, SSH Open Cloud reshapes SSH research's trajectory, amplifying its societal impact.

EOSC-Life (https://www.eosc-life.eu/)
Within the European Open Science Cloud (EOSC), EOSC-Life transforms life sciences. A consortium of research entities collaborates to revolutionize data sharing, analysis, and integration. Dismantling barriers, EOSC-Life empowers life scientists to access, share, and leverage extensive datasets seamlessly. With interoperable tools and data repositories, it accelerates discoveries, from personalized medicine to drug development. Anchored in openness and innovation, EOSC-Life's community-driven approach propels life sciences research and drives global health impacts.

The Research Community

The research community, composed of dedicated scholars, scientists, and experts from various fields, plays a pivotal role in advancing our understanding of complex issues like the COVID-19 pandemic. In this collaborative endeavor, individuals such as Maria Panagiotopoulou, Pieter Fivez, Franciska de Jong, and contributors from institutions like ECRIN and CLARIN ERIC are actively engaged in bridging disciplinary gaps. Their collective efforts harness the power of cross-domain ontologies, harmonized metadata, and AI-driven workflows to enhance the accessibility and comparability of COVID-19-related data. By uniting diverse perspectives and expertise, the research community not only facilitates the seamless integration of data sources but also contributes to more informed decision-making processes for policy development, crisis management, and societal well-being.

The Challenge

The pandemic has led to extensive research activities across various disciplines, such as life sciences, social sciences, and humanities (SSH). However, the lack of interoperable ways to describe these research outcomes has hindered the potential insights that could be gained from combining heterogeneous forms of research. This challenge requires integrating metadata from different sources to facilitate the coupling of COVID-related societal issues with scientific concepts across disciplines.

EOSC service or tool used

The project leverages the EOSC (European Open Science Cloud) as a resource for integrating cross-domain ontology, harmonized metadata, and AI-driven workflows for enhanced research and analysis.

Useful tips & tricks

The project emphasizes the importance of knowing public opinions to effectively communicate policies. It employs harmonized metadata, ontologies, and AI to achieve this goal.

Solution

Harmonized metadata for COVID-related data can enable the coupling of public political debates with biomedical records, societal responses, and opinions from sources like social surveys and social media. An ontology of COVID-related topics has been developed using parliamentary data and social media, allowing for easy linkage to diverse scientific fields. This harmonized model includes links to medical ontologies and identifiers of public policy issues. The project utilized the language-independent representation format developed in the context of ParlaMint.

Across disciplines

The project's impact spans various disciplines, including text mining, humanities, life sciences, and social sciences. The integrated approach ensures that COVID-related research is accessible and understandable across different scientific fields.

Benefits and impact

The developed ontology and workflow enable efficient analysis of public discourse, enhancing understanding of the democratic process, societal tendencies, and public attitudes. This approach extends to social media data and survey data, contributing to effective policy-making and crisis management.

Why do I need EOSC?

EOSC provides a platform for integrating resources like cross-domain ontologies and metadata harmonization tools. It supports data accessibility and interoperability, enabling effective research and decision-making across disciplines.

Useful material related to this story

eoscfuture-grants.eu

rda-alliance.org

Bridging the gap between commercial service providers and researchers through the EOSC ecosystem

Stories from the OCRE/EOSC Future Open Calls

EOSC Future · OCRE

EOSC in practice story #18

Empowering Research Communities through cross-sector data sharing

Enabling Collaboration and Advancements in High-Performance Compute through Open Science Lab

Constanze Roedig

Head of the Austrian DataLAB and the Austrian Open Cloud Community, Vienna Scientific Cluster

"By lowering the entrance barrier, our vision is to make High Performance Computing accessible, usable and approachable by scientists across disciplines, students and the general public. Sharing insights across fields is necessary for truly advancing rather than constantly reinventing the wheel."

Background

As part of an adoption funding programme supported by the EU and building on experience gained through the GÉANT-led OCRE project, the EOSC Future project allotted a total of €4.8 million across three calls, aimed at bridging the gap between commercial service providers and researchers through the EOSC ecosystem.

Specifically, the first call in the series awarded five projects for making commercial services more accessible to Research & Education communities via the EOSC Marketplace, to digital service aggregators (e.g. non-profit entities, NRENs, RIs and e-Infrastructures, HPC centres, etc.) teaming up with OCRE cloud service providers.

The digital service aggregator

The Vienna Scientific Cluster (VSC) provides High-Performance Compute services to all of Austria and acts as a PRACE training centre. To further enhance Research and Research-Driven Teaching, they have begun incubating Cloud Services on top of High-Performance Compute Infrastructures. The Austrian Datalab and Services Project is developing an Austrian wide federated cloud-infrastructure via using automation and DevSecOps, and is maintaining infrastructure blueprints. Sparkle and Technische Universität Wien are using EOSC Future INFRAEOSC-03-2020 Grant Agreement Number 101017536 to make these services available on EOSC.

The user Community

Research groups in traditional High Performance Compute fields such as Physics, Chemistry, Geoscience can use this service to host these interactive Labs for Outreach, Summer schools and Onboarding. Also emerging communities like Machine Learning and DataScience can benefit greatly by using the Analysis-Labs. On the development side, they are also an active research topic for CyberSecurity and OpenScience design principles.

The user Perspective

Master and PhD students can use real HPC results data together with the original tooling (e.g., bespoke coordinate transformation libraries) to experiment with scientific questions or conduct data analysis without being exposed to technical problems like compilers or dependency management.

Why do I need EOSC?

EOSC provides the essential central location to discover, publish and consume the HPC labs and results, which is key for any opensource and/or FAIR project. Furthermore, our intent is the standardisation or, at least, collaborative co-development of common low-level infrastructure services, and for this we require access to EOSC-core like AAI but also an integration with Helpdesk.

The Challenge

The sharing of insights in science and the passing on of knowledge is an ever-increasing issue due to the proliferation of scientific subdomains and the growing specialisation of each sub-field. For scientists themselves, the effort required to stay up-to-date within their own field is growing. Sharing insights across fields is necessary for truly advancing rather than constantly reinventing the wheel. Thus, we require methods aiding this translation and reducing the toil on the scientist, to share, teach, collaborate, peer-review, learn and publish.

Our vision is to use commoditise access to science and to encourage newcomers by lowering the learning curve and cognitive friction in the field of High-Performance Computing.

EOSC Future   OCRE

EOSC in practice story #18

Keywords:
#Interdisciplinary, #OpenScience, #Providers,
#HPC, #OpenScienceLabs, #ResearchCommunity,
#EOSCinPractice #OCRE

The solution

OpenScienceLabs (OSL) enable the sharing of HPC code and results using cloud technology. It provides a universal wrapper for running code in the browser, with access to parallel or special purpose hardware. OSL utilises cloud-native methods and DevSecOps practices, such as containerisation, orchestration, automation, and security measures. It simplifies usage for new or non-expert users by eliminating the need for compilation, installation, or operational tasks. With OSL, end-users can access the system within minutes through a web browser. After an initial co-development phase, any HPC codebase can be made compatible with OSL. The OSL design is continuously improved through use-cases like Peer Review-Companions, Outreach, Open Sourcing, Self-Service On-boarding, and HPC-Summer Schools. These use-cases help enhance usability, scalability, and portability of OSL.

In depth

The "OpenScienceLabs for HPC" are based on open-source blueprints available here: https://dev.azure.com/AOCC/OpenScienceLabs
We implemented an opinionated, hardened, but lightweight Kubernetes stack combined with a generic wrapping mechanism (using JupyterHub) and modern Identity Federation. Our approach focuses on separating state and run-time, enhancing security and enabling independent development of Labs and user data.

The impact on society

Our vision is to democratise High Performance Computing by making it more accessible and usable for scientists, students, and the general public. We aim to lower the entrance barrier, relieve researchers from technical burdens, simplify peer-review processes, and establish a digitally sovereign infrastructure in academic institutions. Through "Open Research Platforms in the Cloud," universities can collaborate using our secure, affordable, and tested OpenSource blueprints. This collaborative effort will improve cost-efficiency, save researchers time, democratise HPC knowledge, and encourage serendipitous discoveries.

Across disciplines

The OpenScienceLabs for HPC is most obviously encouraging the cross-disciplinary research between data science, machine learning and STEM subjects and was built on that premise. Furthermore, research in computer science and embedded engineering will likely be combined in order to improve the security and accessibility.

Sustainability for an EOSC in practice

Most Open Science Labs for HPC are initially development-heavy for a Research Group. However, once established, we plan on offering workshops and blueprints to maintain the operations in their own cloud-native infrastructures. This may require establishing those at various HPC facilities. We are conducting active research on parametrised operational models to quickly estimate the financial feasibility of a use case.

Future developments & funding scenarios

After completing the 3D rendering features, we plan to integrate Machine Learning Workflow tools like Kubeflow/Elyra. This will enable connecting workflow steps to HPC clusters, enhancing security measures, and improving the self-service aspect of our offering. We also aim to provide more blueprints and initial training to Research Groups, although we are currently in the prototype stage.
For future funding models, we are considering shifting OpenScienceLabs more towards teaching. This would reduce operational costs and cloud spending while maintaining the production of open-source artifacts. It could also inspire students to pursue R&D careers. Additionally, we are exploring the possibility of a run-time federation with existing HPC infrastructures, although this poses technical challenges.

Useful material related to this story

ocre-project.eu   open-science-labs

EOSC Future   OCRE

EOSC in practice story #18

Keywords:
#GWDG, #TrustedResearchEnvironment, #TRE,
#AWS, #ResearchCollaboration, #IaC,
#Sustainability, #EOSCinPractice

GWDG Trusted Research Environment (TRE): Empowering European Research through AWS Cloud Services

Overcoming Challenges and Enhancing Collaboration in the European Research and Education Community

Piotr Kasprzak

Cloud Engineer at GWDG -
Gesellschaft für wissenschaftliche
Datenverarbeitung mbH Göttingen

"By lowering the entrance barrier, our vision is to make High Performance Computing accessible, usable and approachable by scientists across disciplines, students and the general public. Sharing insights across fields is necessary for truly advancing rather than constantly reinventing the wheel."

Background

A total of €4.8 million has been allotted to bring ready-to-use, state-of-the-art commercial services to EOSC users. The adoption funding is spread across 3 calls, each on a different theme, in order to boost supplier engagement, build on experience gained through the GÉANT-led OCRE project, offer guidance on governance and compliance, and ensure funding reaches across different domains and scientific communities.

The service provider

GWDG
Gesellschaft für wissenschaftliche
Datenverarbeitung mbH Göttingen

The GWDG is an well known IT service provider and computing center not only for both the University of Göttingen and the Max Planck Society, but also for the broader R&E community in Germany. It is a national HPC centre, a national AI service centre, and a core player in the national German Research Data Initiative (NFDI).The project "GWDG Trusted Research Environment (TRE)" has been funded under "EOSC Future Call 1: European Research E-Infrastructures – Distribution of Commercial Cloud Services" (Grant Agreement number 101017536).

The user Community

The service is being offered to a wide range of research communities supported by GWDG, with particular emphasis on the Life Sciences community with its often stringent data security and regulatory compliance requirements. Within this community, there are already two prominent use cases that have emerged as flagships. Furthermore, use cases from the astronomy community are being evaluated for deployment on the platform, and there is an expectation that additional communities will embrace this infrastructure in the future, given its broad applicability and appeal.

Why do I need EOSC?

By providing a single point of entry for accessing data and services EOSC streamlines the discovery and utilization of pertinent resources and contributes to the essential visibility and recognition required to draw users from various European research communities to the offered service. Furthermore, the funding provided has been critical for the development of this service and onboarding first research projects.

The Challenge

The adoption of commercial cloud services in the European R&E community faces numerous challenges related to operations, security and accessibility. These issues currently prevent researchers from enjoying the well-documented advantages of these cloud services including rapid scalability, flexibility and cost savings. Researchers typically lack the technical expertise required to set up and manage multiple interconnected cloud services to address their research needs. Data security and compliance concerns are particularly difficult to address and are often preventing the use of cloud services in the first place. What researchers require are preconfigured, user-friendly environments with clearly defined workflows that take governance and regulatory processes into consideration. These environments would simplify the consumption of cloud services immensely and thus, in turn, would enable researchers to focus on their research instead of grappling with technical intricacies.

The Solution

The GWDG TRE project tackles the previously mentioned challenges by offering secure and user-friendly environments on the AWS Cloud, specifically designed for collaborative scientists conducting data analysis. These environments comprehensively integrate necessary AWS services, abstracting away technical complexities.

EOSC Future

OCRE

EOSC in practice story #18

🏷 **Keywords:**
#GWDG, #TrustedResearchEnvironment, #TRE, #AWS, #ResearchCollaboration, #IaC, #Sustainability, #EOSCinPractice

To facilitate easy utilization, they feature a customized web-based interface built on AWS Service Workbench as the primary entry point.Addressing data security and complex compliance needs, including those spanning multiple legal jurisdictions, is achieved through the implementation of a governance process with various roles and rigorous control over data ingress and egress, whenever necessary. The project relies on an AWS Landing Zone design as its foundational technical framework, which is extended by an AWS Trusted Research Environment (TRE) implementation tailored to the specific requirements of the R&E community.In addition to these technical components, the project also conducts outreach activities such as AWS Research Immersion Days, complementing the overall target to facilitate the distribution of cloud services in the R&E communnities.

Useful tips & tricks

The service has been developed using the latest Infrastructure-as-Code (IaC) frameworks for AWS, notably AWS CloudFormation and the AWS Cloud Development Kit (CDK). The latter enables the use of widely-used programming languages for modeling, with automatically generated AWS CloudFormation templates for deployment. The standard GitOps workflow is employed for both development and deployment, which seamlessly aligns with the mentioned technologies and the broader automation capabilities offered by AWS.

The impact on society

Reducing the obstacles that hinder the adoption of commercial cloud services within the R&E communities holds the promise of expediting crucial research endeavors. It opens the door to vast, scalable, elastic, and cost-effective computing resources, which have the potential to accelerate essential research and provide solutions to a wide array of contemporary societal challenges. These challenges span from the discovery of new or enhanced vaccines and treatments for diseases to addressing the complex demands of transitioning towards a more environmentally sustainable economy and mitigating the climate crisis.The impact of synergizing the brilliant minds in diverse research communities with the nearly limitless processing and analytical capabilities, as well as the innovative services offered by commercial cloud providers like AWS, is indeed difficult to overstate.

Useful material related to this story

Across disciplines

The TRE environment furnishes a complete set of tools and procedures for seamless integration and accessibility of data analysis tools, catering to researchers from various research fields. This ensures that interdisciplinary research collaborations benefit from a robust technical underpinning to collectively address research challenges. TRE also enables the establishment of workflows tailored to discipline-specific regulatory or operational standards. These principles hold true for cross-national research initiatives as well.

Sustainability for an EOSC in practice

GWDG TRE is being implemented on top of an preexisting AWS Landing Zone, which has been funded and supplied by GWDG. GWDG views TRE as an essential component of its AWS service offering for the R&E community. Consequently, GWDG plans to maintain funding for the required AWS infrastructural ressources, as well as invest in further development internally thereby guaranteeing sustainability.

Future developments & funding scenarios

The project's funding and relatively short timeline allow for the initial implementation of essential features within the TRE environments, primarily focusing on critical AWS services for data analytics. Future development of the TRE environments will progress along three main avenues:Expansion of Plugins: The list of additional plugins to support diverse requirements, special customisations or certain functional expansions is continually growing as the AWS TRE toolkit evolves and expands further. These can be integrated into the GWDG TRE environment depending on concrete requirements and available funding.Tailored Services and Workflows: For specific scientific use cases more customized, pre-configured services can be offered within the environment. This can also involve the development of corresponding supplementary workflows.Integration of more specialized AWS Services: The general capabilities of the environment can be further expanded by the integration of additional, more specific AWS services. This will become more important in the future with the expected advent of specialized AI services.In essence, the current project marks only the beginning, with extensive ongoing development and expansion planned along these mentioned paths in the future.

GWDG

QCloud - Ireland's quantum cloud environment

Background

As part of an adoption funding programme supported by the EU and building on experience gained through the GÉANT-led OCRE project, the EOSC Future project allotted a total of €4.8 million across three calls, aimed at bridging the gap between commercial service providers and researchers through the EOSC ecosystem.

The third and last call in the series awarded two projects for making commercial services more accessible to Research & Education communities via the EOSC Marketplace, to digital service aggregators (e.g. non-profit entities, NRENs, RIs and e-Infrastructures, HPC centres, etc.) teaming up with OCRE cloud service providers.

Hazel Murray

PhD and Cybersecurity Lecturer at MTU – Munster Technological University

"The AWS infrastructure and the TRE developed as part of this project will not only improve the quality of Irish quantum research but will also provide ease of collaboration between national and international distributed research teams, promoting cross border collaboration and resource sharing."

Dr Hazel Murray MTU

Overview

MTU — Olloscoil Teicneolaíochta na Mumhan / Munster Technological University

Harnessing the power of quantum mechanics, this project offers cloud-based access to quantum computing hardware. This breakthrough technology opens avenues for cryptography, networking, data analysis, and optimisation. Researchers gain access to a variety of quantum processors, driving experimentation and innovation. Harnessing quantum computing through the cloud, enables researchers scattered across Ireland to access and leverage this cutting-edge technology, fostering collaboration and accelerating scientific progress.

"Access to quantum computing not only empowers us to personally unveil the advantages of quantum computing in the field of security of autonomous systems and lay the foundation for a quantum-powered future but also enables us to actively engage in software engineering for quantum computing, all while nurturing the minds of the next generation of students and innovators in the realm of quantum knowledge"
Dr Anila Mjeda, CyberSkills, MTU

Quantum computing infrastructure demands exceptionally low temperatures, precise calibration, and expert maintenance, making it a substantial financial and logistical burden for individual researchers or smaller institutions. By providing cloud access, researchers can tap into these cutting-edge quantum resources, democratizing access and fostering innovation in the field of quantum computing. This accessibility not only accelerates scientific discoveries but also fosters collaboration and knowledge sharing among a wider community of researchers, ultimately advancing the frontier of quantum computing research.

This EOSC funding will allow Ireland-based researchers to gain access to public cloud quantum computers through AWS, and support their research. Currently, researchers rely on free introductory credits to run small-scale experiments on the private cloud or on quantum computing simulators. This distribution of cloud access will support greater ambitions within the projects themselves and help familiarise researchers with the commercial services available to support ambitious projects.

The Challenge

The Current Entrance Barrier to Quantum Computing Technology

Quantum computing is a disruptive technology that harnesses the laws of quantum mechanics to solve problems outside the scope of our current classical computers. Quantum computers will be a core component of future networks, offering capabilities in cryptography, networking, data analysis and optimization. However, a quantum computer is difficult and expensive to build.

The solution & the commercial cloud service

Quantum cloud access for Irish based researchers.

The central platform in which the aggregator, MTU, will promote and stimulate the distribution of Cloud Services within the research community in Ireland, is based on an AWS solution called Trusted Research Environment (TRE). This is a self-service research solution to help secure and analyse data in a collaborative manner, whilst delivering a set of flexible tools and pre-built templates to conduct research at scale. MTU will use the TRE as the centralising mechanism in which it will distribute Quantum Computing capabilities amongst its partners and the wider research community in Ireland.

The ultimate goal is that any researcher in Ireland with Quantum Computing needs will be able to reach out to MTU and make their case to use the TRE as the platform for their research. This way, MTU aims to facilitate Quantum Computing access to the wider Irish research community.

Based on the advantages of basing quantum computing access on a cloud-based system, MTU will work with researchers across the country to stimulate the distribution of quantum cloud computing. Currently, researchers from 8 Irish universities and 13 research institutes have identified an initial pipeline of nine proposals. The shared goal of these researchers is to combine quantum and classical networking and computing to create the quantum internet and solve the problems of today with the quantum technology of tomorrow.

"Quantum-computer simulations for quantum chemistry can, in principle, yield exact solutions to complex many-body problems. This will enable us to solve materials science problems at a level of accuracy not in reach with standard computers."

Prof. Stefano Sanvito Professor of Condensed Matter Theory, Director of CRANN

The projects funded by this call will have a direct impact on individuals, businesses and governments who rely on secure communication and transactions; on AI, autonomous vehicles, adaptive living, smart tech and assistive living; on drugs and procedures developed by the pharmaceutical and medical professions and on our understanding of the natural world including predicting the progress and effects of climate change.

As Quantum continues to emerge and increases in importance, it will become essential for students and researchers to have access to quantum computing resources. The scalability of the public cloud ensures that as their field grows in importance and as we move towards a hybrid classical/quantum era, access for all Irish researchers and educators to quantum computing will be feasible. This EOSC funded quantum cloud environment is essential to ensuring the quality and accessibility of this cutting-edge resource, keeping Ireland at the forefront of technological progress, and enabling researchers and developers to test and create quantum algorithms of the future.

IMPACT

Quantum computing is the future of technology

Quantum computing is the technology of the future. It can provide secure communication and data exchange via entangled qubits, it can break our current modern cryptography, it can speed up classical optimization problems allowing faster AI/ML and Data analysis and it can allow us, for the first time, to truly simulate events from the real world that include a degree of randomness. It is a powerful tool which will provide a deeper understanding of the world around us and a strong security system for communications.

Useful material related to this story

MTU

Supporting companies to access EOSC digital technologies and services

Stories from the EOSC DIH/EOSC Future Open Calls

EOSC Future

Pundit: Empowering Web Annotation for Researchers and Beyond

Enhancing Information Gathering, Organization, and Collaboration

Luca De Santis

Chief Technology Officer at NetSeven

"Pundit, developed by the Italian company Net7 since 2011, is a web annotation tool that simplifies the process of 'taking notes' on web documents, offering structured and reusable annotations for researchers, scholars, and everyday users."

The Project Involved

The Italian company Net7 (https://netseven.it) has been developing, since 2011, the web annotation tool Pundit (https://thepund.it). Pundit helps users, in particular researchers and scholars, to create annotations, that is, to easily and effectively "take notes" on web documents like HTML pages and PDF files.

The Challenge

Many people on the web, including researchers, students, and everyday users, often face three key challenges. They struggle to:
- Gather important information from the web by taking notes.
- Find a single place to keep and manage these notes.
- Make these notes more structured and easy to reuse.

These challenges aren't limited to experts; they affect anyone trying to make the most of online information. So, whether you're a student doing research or just someone looking for information online, these difficulties are something many of us encounter.

EOSC service or tool used

Pundit can be an answer to these issues. It is a cloud service that allows users to "take notes" on web documents, like a web page or a PDF file. It consists of a set of components, amongst them the Annotator, a free extension for the Google Chrome browser used by users to create annotations. Annotations can be highlights of text parts, comments, tags but also semantic annotations, real RDF triples (subject -> predicate -> object) to specify formal, structured and reusable statements associated to the annotated content.

Pundit can be found on the EOSC: https://marketplace.eosc-portal.eu/services/pundit-a-semantic-powered-web-annotation-service

Useful tips & tricks

Pundit is free to use. You can go to its web site https://thepund.it, register, download the Chrome extension and start to annotate the web. Your annotations will be available in the the Pundit web app and can be easily exported in open formats, to encourage their maximum reuse.

The Research Community

Currently Pundit has almost 20.000 registered users, mostly scholars, especially those belonging to the SSH domain. The 100 most active users in the last 12 months created about 13.000 annotations.
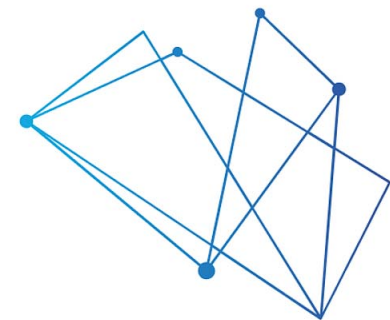
Benefits and impact

The inclusion of Pundit in the EOSC marketplace has had a profound impact on its visibility and potential utilization. Previously, Pundit was primarily recognized within the SSH (Social Sciences and Humanities) domain. However, its presence in the EOSC marketplace has broadened its reach, making it accessible and relevant to a wider audience beyond SSH. Additionally, active involvement in the EOSC DIH pilot initiative has opened new horizons. It has prompted a thorough exploration of alternative business models and diverse use cases for Pundit. These efforts aim to unlock its potential in areas such as training artificial intelligence models for text processing, transcending its original scope and contributing to a more versatile and extensive adoption of the tool.

Why do I need EOSC?

EOSC is and will be even more significantly in the future, the one-stop-shop for European researchers in search of solutions to solve their problems and to improve the way they work. Being there is therefore essential for a solution like Pundit, which has been mostly implemented through EU funded research projects and has been designed from day one to support research activities.

Across disciplines

Albeit Pundit is totally agnostic respect the scientific domain in which it is applied, it is also true that historically the SSH domain has been the most receptive in using a tool "to take notes on the web". SSH is itself a wide and cross-disciplinary domain, with at least 27 branches of knowledge, spanning from Archeology to Economy, from Social Sciences to Art and Art History. Finally Pundit encourages the sharing of annotations amongst users, facilitating collaborations, possibly also in a multidisciplinary way.

Limitations and future improvement

Pundit is continuously growing with new features. One area that we are actively working on is to potentiate the support of Semantic Web features, allowing users to annotate by using multiple ontologies and also by defining their own personal ontologies.

Useful material related to this story

netseven.it

thepund.it

Stories from the EOSC Community

Increasing the service offer of the EOSC Portal

Stories from projects funded under call INFRAEOSC-07-2020

EOSC Future

Mapping Mesophotic Habitats with EOSC Services and Tools
Enhancing Marine Habitat Mapping through Remote Sensing and Research Object Preservation

The project involved

The EU-funded RELIANCE project seeks to extend the capabilities of the EOSC with enhanced support for a range of research activities aligned with the EOSC Interoperability Framework. This will make it easier to find and use research data, including Copernicus data, simplify extraction of relevant information, and effectively managing the research lifecycle via research objects. The project also supports the principles of FAIR and open science.

Reliance

"Marine habitat mapping aims at providing a holistic representation of ecologically relevant marine habitats, their associated biological communities, their areal extent, distribution patterns, status, and physical conditions."

Giorgio Castellan, Postdoctoral Researcher at CNR-ISMAR Institute of Marine Sciences, RELIANCE

The Research Community

The marine habitat mapping community brings together ecologists, biologists, and geologists to provide accurate and extensive maps of marine habitats, the community aims at detecting environmental changes in marine habitats using an ecological approach to support marine spatial planning, to help reduce the cost of monitoring programmes and provide environmental impact assessments.

The Challenge

Marine habitat mapping provides a holistic representation of ecologically-relevant marine habitats, their associated biological communities, areal extent, distribution patterns, status, and physical conditions. The mesophotic zone, part of the water column with dim light conditions, can host a diverse mosaic of benthic habitats, hotspots of biodiversity. However, the mesophotic domain depends on light conditions that vary considerably depending on geographical location. Satellite and remote sensing data can be used to map the extension and distribution of the mesophotic zone, providing information to support the conservation and management of the biodiversity the zone hosts. Integrating and processing satellite/remote sensing data, however, can be challenging in terms of data mining, data accessibility and computational resources. This was one of the main challenges of the project.

EOSC service or tool used

A Python script was used to model the mesophotic zone using EGI Jupyter Notebook (notebooks.egi.eu). Satellite data on Photosynthetically Active Radiation (PAR) and diffuse attenuation coefficient for light at 490 nm were then obtained from the ADAM Platform (reliance.adamplatform.eu) and used to map the Mediterranean mesophotic zone. A Research Object was created in ROHub (reliance.rohub.org) containing the inputs and outputs of the model, and the Jupyter Notebook running the model. As this is fully public and accessible, the Research Object can not only be included in future research, but it is also findable on OpenAIRE Explore (explore.openaire.eu) and Zenodo (zenodo.org).

Useful tips & tricks

The ADAM Platform (reliance.adamplatform.eu) allows the extraction of global and local remote-sensing data from past and current times, as well as short-term forecast and long-term projections. By collecting data from different sources, such as Copernicus, NASA Ocean Color, etc., ADAM makes remote-sensing data from different repositories easily-findable and accessible, from the web platform and from Jupyter Notebook (ADAM API).

ROHub (reliance.rohub.org) is a holistic solution for the storage, lifecycle management, and preservation of scientific investigations via research objects. It makes these resources available to others, allows publishing and releasing them through a DOI, also through Zenodo, allowing to discover and reuse of existing scientific knowledge. The Research Objects can also be accessed from EOSC OpenAIRE Explore.

Benefits and impact

EOSC services provided a set of tools that ensured the research process followed FAIR principles. The Jupyter Notebook allowed the model to run in a collaborative and powerful environment and the process to be shared with other researchers. The interoperability with ADAM Platform ensured the exploration and access of a large volume of remote-sensing data, used as inputs for the model. Further, the encapsulation of the research process in a Research Object allows our research to be easily accessible to other researchers, boosting the impact of the research itself. The reproducibility of scientific approaches, sharing methods, and results are of primary relevance for the scientific community, allowing to capitalize on available information to increase the knowledge of our oceans.

Why do I need EOSC?

"*Within the Habitat Mapping community, I used to run my models and work on local environments and share the results through direct contacts*" states Giorgio Castellan, Postdoctoral Researcher at CNR. "*However, scientific research is moving fast towards a more collaborative approach,* *both in terms of methods and exploration and capitalization of existing data. The EOSC services simplified the collection of existing information on marine habitats and remote-sensing data, strongly reducing the amount of time dedicated to this task. Further, the EOSC tools made my research activities to be more reproducible and sharable with other experts. I found very useful the set of services developed and included in the EOSC portfolio related to remote-sensing.*"

Across disciplines

Marine habitat mapping is based on the integration of ecological, biological, and geological aspects to provide accurate and extensive maps of marine habitats. In the past, searching for information on different research themes required a huge amount of time, often relying upon personal contacts. The EOSC Portfolio of helps overcome this barrier, allowing different disciplines and expertise to combine in a powerful and modern environment.

Useful material related to this story

» https://www.reliance-project.eu/

Want to learn more about the other services being developed by **RELIANCE**? Read here
Liked this **#EOSCinPractice** story? Follow @EOSCFuture for more!

EOSC Future

ROHub: A Resource Organisation Service for Scientific Investigations

Facilitating Reproducibility, Reuse, and Citation of Research Materials and Methods

Elisa Trasatti

Senior Geophysical Researcher
Istituto Nazionale di Geofisica e
Vulcanologia, Italy – RELIANCE

"The RO service helps to organize resources, materials, and methods of an investigation. Share research materials with other scientists, facilitate reproducibility and reuse of scientific methods. Be recognized and cited (uniquely identified by an URI, or a DOI). Preserve results and prevent decay. Provide evidence to findings claimed in scholarly articles."

The Project Involved

Reliance

Reliance (Research Lifecycle Management Technologies for Earth Science Communities and Copernicus Users in EOSC) is a H2020 project funded with Grant Agreement No.101017501. The objective of the project is to extend the research-enabling capabilities of EOSC with services for research lifecycle: the Research Objects (ROs) as the overarching mechanism to manage scientific research resources, the data-cubes for scalable structured data access and discovery, and text mining services to extract machine-readable metadata from ROs.

The Research Community

Geohazard is a worldwide research community devoted to the study and management of natural hazards such as volcanic eruptions and earthquakes. Working in this domain implies field work to install instruments, the use of Earth Observation monitoring systems, simulation of potential hazards and direct interaction with decision makers and stakeholders.

The Challenge

Geohazard community is a global community dealing with scientific problems with fragmentation, unequal access to data and research methods, and isolation. This is critical as volcanic eruptions or similar natural disasters can impact neighbouring countries and even entire continents. The main challenge faced by Geohazard therefore is to create services that can ease access to data and tools, conduct high-quality research, and share the findings. These services should provide timely and scientifically accurate information about volcanic and seismic activity, not only within the research community but also to various end-users and stakeholders at national, European, and international levels.

EOSC service or tool used

There are several services available via EOSC that can be used for the challenges mentioned. One example is the EGI notebooks, used through the Reliance Virtual Organization, which are useful to access data, to run tools, and share data, methods and results in the cloud with other colleagues. The RO service provided through the ROHub portal is also extremely significant to helps organise and describe resources, materials, and the methods of an investigation. In particular with ROHub it is possible to:

- Share research materials with other scientists and end-users;
- Facilitate reproducibility and reuse of scientific methods;
- Be recognized and cited (uniquely identified by an URI, or a DOI);
- Preserve results and prevent decay;
- Provide evidence to findings claimed in scholarly articles.

The service has been leveraged to the EOSC marketplace during the H2020 Reliance project. https://tinyurl.com/5h4u8mfr

Useful tips & tricks

The RO paradigm can be used to document research activity in several ways: to share Earth Observation data products (e.g., the Changbaishan volcano1, China/North Korea, surface deformations), to timely share reports and materials (e.g., during the volcanic emergency of the Nyiragongo volcano2, DR Congo), and to fully document a published tool (VSM, Volcanic and Seismic source Modelling3 with links to the code and to examples, the latter being themselves ROs).

1. https://doi.org/10.24424/vfp6-r230
2. https://w3id.org/ro-id/45841548-0362-4aea-80f2-ea71d81a691f
3. https://doi.org/10.24424/t83f-5t97

EOSC Future

Benefits and impact

During the 21-22 May 2021 eruption of the Nyragongo volcano (DR Congo), my team and I have been promptly working analyzing Earth Observation data (Sentinel-1) to map surface deformations, and running modelling tools to undisclosed the position and depth of the magma feeding the eruption. The RO service in the EOSC allowed me to rapidly share our scientific information not only among the Geohazard global community, but also to international end-users and stakeholders, confident that, even in the midst of an emergency, IPR was properly attributed. The benefit of using the EOSC service is not only cross-disciplinary, but goes beyond the scientific community domain.

Why do I need EOSC?

EOSC is an infrastructure that can be used by researchers and practitioners of any scientific discipline. There is a large suite of services available that can improve how the researcher is carried out, and therefore boost the results sharing. For example, during the Reliance project it has been possible both to use a EOSC service (the EGI notebooks and cloud), and to implement new services, such as the ROs, the data-cube for Earth Observation data access and discovery, and the text-mining service.

## Across disciplines

By definition, the RO service is used to document any type of resource of the research lifecycle, therefore it can be widely used in any domain and is cross-disciplinary by nature. In addition, among the different tools available to document the RO, it is possible to create relations among the current RO and others.

Useful material related to this story
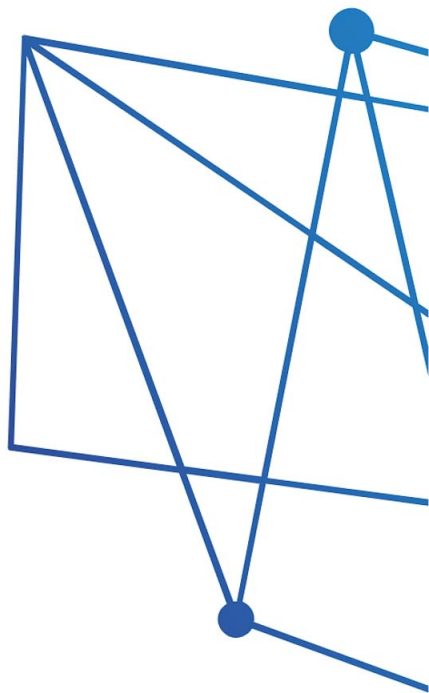
reliance-project.eu

EOSC Future

Advancing Nordic Climate Science with Innovative EOSC Services and Tools

Enabling Nordic climate scientists to overcome collaboration challenges and accelerate research with EOSC services and tools

Anne Fouilloux

Anne Fouilloux, Senior Research Engineer at Simula Research Laboratory, Norway - RELIANCE

"The benefits from using EOSC services and tools for individual researchers and small communities are immediate: it is impossible to achieve such a level of integration in a small laboratory, especially when you work in a "small" country such as Norway where skilled professionals are scarce."

The Project Involved

Reliance

RELIANCE (Research Lifecycle Management Technologies for Earth Science Communities and Copernicus Users in EOSC) is a H2020 project funded under GA n.101017501. The project extends the research-enabling capabilities of the EOSC with services for research lifecycle: 1) Research Objects (ROs) as the overarching mechanism to manage scientific research resources, 2) data-cubes for scalable structured data access and discovery, and 3) text mining services to extract machine-readable metadata from ROs.

RELIANCE services are co-designed with Earth Science Communities and in particular by the climate science community, represented in the project by the University of Oslo (NO) and Simula Research Laboratory (NO). Here research efforts were devoted to analysing the impact of covid-19 lockdown on air quality, and on the impact of extreme events on arctic vegetation and potential threats for local populations.

The Research Community

The Nordic climate science community comprises researchers, software engineers and students covering a broad range of relative fields such as meteorology, chemistry, biology, ecology, and, more in general, anyone with an interest in environmental data science.

The Challenge

Climate science is multidisciplinary by nature as it involves research in diverse disciplines such as physics, chemistry, biology, geology, geography, meteorology, oceanography, etc. Common tools, data and techniques used for climate include numerical models, big data analysis, AI, in-situ observations, remote sensing and laboratory experiments.

To handle the huge, ever growing and distributed amounts of data involved, the software stack is diverse and complex. Also, given that the tools available are written in different programming languages (Python, Julia, R, Fortran, C, C++) and that scientists often possess different backgrounds and technical skills, this only adds to the challenge of multidisciplinary collaboration. Open Science is undoubtedly the most effective way to overcome these difficulties and EOSC provides the framework to put it into practice.

EOSC service or tool used

Anne Fouilloux, Senior Research Engineer at Simula Research Laboratory, Norway - RELIANCE says:
"For us it wasn't so much about using one EOSC service or tool but the integration, communities and people. "

Sharing and collaboration are made easier with the same computational environment, tools and data; avoiding "it works for me, why does it not work for you?" scenario.

We started our EOSC journey with RoHub, aggregating artefacts into "Research Objects", with the ADAM platform for large datasets access and visualisation, and EGI Notebooks for data analysis. When storage and computational resources became problematic, we got access to larger JupyterHubs through EGI-ACE and C-SCALE projects.

EOSC is an endless journey where practitioners thrive and push the boundaries of science and technology.

Useful tips & tricks

As researchers, if you think that you cannot implement EOSC services and tools: you are wrong! In the RELIANCE project (https://www.reliance-project.eu/), RoHub (https://reliance.rohub.org/), the ADAM platform (http://reliance.adamplatform.eu), the Text-mining services (https://reliance.expertcustomers.ai/) are co-designed. This is usually true for all the EOSC services and tools and your feedback is more than welcome!

EOSC integrates many services and tools, which can seem overwhelming at the beginning. Authenticating to EOSC services and tools with your ORCID identifier is the best approach in our opinion.

A good approach for newcomers would be to identify the biggest bottleneck in their day-to-day work and find the EOSC service that could overcome it: start simple, small and then grow.

Bear in mind that those who develop and deploy services and tools are not end-users: be gentle with yourself and others. Finding another researcher using EOSC, even if working on different scientific fields, is usually the way to go.

Benefits and impact

The benefits from using EOSC services and tools for individual researchers and small communities are immediate: it is impossible to achieve such a level of integration in a small laboratory, especially when you work in a "small" country such as Norway where skilled professionals are scarce. Before, we were wasting a lot of time moving large amounts of data from one site to another (HPC to cloud to laptop). Access to ARCO (Analysis Ready Cloud Optimized) data through easy-to-use compute and storage resources such as Jupyter Notebooks significantly improved our daily routine.

Collaboration is fast and easy: data, compute and storage are always there and accessible to all. Using Research Objects (ROs) to aggregate everything we use and produce (data, software, videos, documents, etc.) changed our mindset and sped up our work. Rather than writing irrelevant Data Management Plans (DMPs), we used RO as our DMP and created regular snapshots to capture the entire life cycle of our research project. When the time to publish comes, we are ready and can share the ROs in our paper.

We are more visible (our research is more re-used and not only because it is fully reproducible), have more (and more diverse) collaborators and access to more efficient services and tools, and we also learned to be mindful when sharing (FAIR and CARE principles).
Finally, we understand the role we can play to shape EOSC and benefit from it.

Why do I need EOSC?

EOSC plays an important role for our small community: we can push the limits of the possible and as small as we are we can think big! There is no way back: we cannot imagine having to download, prepare data and the needed computational environments locally and by ourselves, and the time involved in trying to share our work with all our collaborators. For example, when we needed larger amounts of compute & storage to run machine learning algorithms to study the impact of extreme events on arctic vegetation (see RO titled "Vegetation browning in Troms and Finnmark (Norway)" (https://w3id.org/ro-id/3ed30e69-fb38-4045-bd34-2fa907d12353), a customised deployment of the Pangeo (https://pangeo.io/) machine learning notebook (https://github.com/pangeo-data/pangeo-docker-images) was made available through collaboration with EGI-ACE and C-SCALE.

On-boarding new community members is also easier. For instance, students do not need to reinvent the wheel: we point them to relevant executable Research Objects they reproduce and build upon. The learning curve is smoother. The RO titled "OCTOPUS project - explOring aerosol-Cloud inTeractiOns in CMIP6 models Using joint-histogramS" (https://w3id.org/ro-id/dd948b04-bfa4-44b0-814b-19f7daff6b8c) by Adele Zaini is a good example.

Another clear benefit from EOSC is when you get a new job at another institution (as many researchers do): you can still be authenticated to the same services and tools with your unique ORCID identifier and continue to collaborate and work.

Useful material related to this story
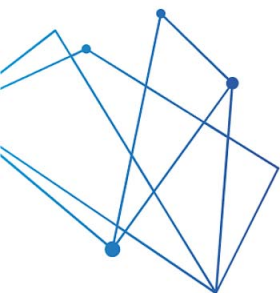
reliance-project.eu

Across disciplines

Working across disciplines is also facilitated by EOSC services and tools. Here is an example: our climate community focused on the impact of the lockdown on air quality (https://w3id.org/ro-id/53aa90bf-c593-4e6d-923f- d4711ac4b0e1) while the Sea Monitoring community from RELIANCE did a similar study but analysed the water quality (https://w3id.org/ro-id/0869e396-3733-4aff-8fb2-94c8937b28aa).

The two original ROs are fully reproducible and reusable so it was fairly easy to combine the two studies to derive a new Jupyter notebook as demonstrated in the executable RO titled "Changes in air quality and water quality during the Covid-19 Lockdown in the Venice Lagoon (https://w3id.org/ro-id/ 998dccd6-7192-4d88-af39-6018c71e6bdf).

Open Science is a process and thanks to EOSC, we understood that making our research accessible to everyone is not necessarily difficult but it is very different from what we were told: it is a lot about people and communities of practice play an important role.

For instance, agreeing on and following clear guidelines for writing modular and reusable Jupyter Notebooks make executable Research Objects much more reusable. This is why we have teamed up with the Environmental Data Science book (EDS book, https://edsbook.org), a pan-european community-driven resource that enables researchers to co-design, collaboratively and openly review, and curate interactive, shareable and reproducible executable notebooks.

Limitations and future improvement

Tools and services are not meant to be static: they need to evolve with researchers' needs. It should not be any different for EOSC but we feel that the sustainability of the EOSC services and tools are not always clear.

Engaging and on-boarding researchers on EOSC require long-term sustainability of the tools and services provided, taking into account that future needs will emerge and will have to be taken care of.

Interoperability of the EOSC services and tools outside Europe is also important: the goal is not to create a bigger silo and isolate European researchers that can use EOSC from others. Open Science cannot be realised with Europe only. We suggest dropping the E and think about a new chapter e.g. OSC "Open Science Cloud"!

Useful material related to this story

reliance-project.eu

Athena RC: Empowering Open Science and Digital Transformation for Rural Communities

Leveraging Gnomee web application and interdisciplinary solutions to drive innovation in Greece

Panagiota Koltsida

Research Associate at National and Kapodistrian University & Athena Research Center

"EOSC offers an integrated research environment that makes services such as OpenAIRE AAI available and accessible. It serves as the single access point for identifying relevant services that suit a specific context, purpose and/ or new service".

The Project Involved

Research & Innovation Information Technologies / ATHENA

Athena RC is a research organisation focusing on designing and developing inter-disciplinary solutions for research communities and society. It is one of the main Open Science ambassadors in Greece and supports EOSC through various projects. As partner in the DESIRA Horizon 2020 research project (https://desira2020.eu/), Athena RC has designed and developed a set of digital tools to facilitate digitalisation in rural areas and agriculture, and strengthen the impact of Living Labs.

The Research Community

Gnomee is a web knowledge base that contains Digital Solutions which can act as potential Digital Game Changers in rural areas, agriculture, and forestry. Each digital solution is classified under a set of taxonomies useful to farmers, facilitators and researchers willing to identify solutions for these areas for exploitation or further enhancement.

The Challenge

To ensure wide acceptance and exploitation of digital tools from - but not only- the research community, it is important to re-use and exploit common solutions and services with regards to core functionalities and operations and not introduce additional options, compared to existing ones. This approach should be followed from applications that require users to create an account. The challenge, in this specific context, is to eliminate the need for asking users to create yet another account and increase the possibility of exploiting the full potential of your web application.

EOSC service or tool used

To solve the above challenge, we exploited OpenAIRE's AAI service to offer external authentication through a list of well-known providers, including academic institutions, ORCID, as well as social providers like Google and Linkedin. Using this service, users can login to our application through their existing accounts and can take advantage of the functionality offered to authenticated users.

OpenAIRE's AAI service offers a standard integration method and can be easily integrated by external web applications.

Useful tips & tricks

The main authentication provider for our web application: Gnomee (https://www.gnomee.eu) is OpenAIRE AAI service allowing our potential users to be easily authenticated and become members of our application and exploit it further. Integrating with OpenAIRE AAI is an easy task as it supports OAuth2 and SAML standard protocols.
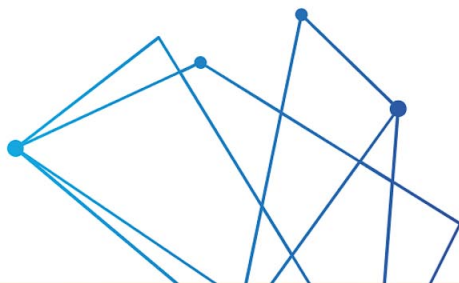
EOSC Future

Benefits and impact

Gnomee Knowledge Base is a community driven tool and, as such, offers the functionality to be updated with new Digital Solutions by users. To support this feature, users need to login to the application in order to suggest new solutions to become part of Gnomee. This feature has been easily implemented exploiting OpenAIRE's AAI EOSC service as an initial step for authenticating the user without requiring significant effort by the user itself allowing him/her to focus on the information of the new suggested digital solution.

Ensuring that Gnomee is updated with new digital solutions, increases its sustainability and potential impact, and facilitates its continuous exploitation from the targeted users.

Why do I need EOSC?

EOSC offers an integrated research environment that makes services as the OpenAIRE AAI available and accessible. It acts as the single point of access for identifying interested services, suitable for a specific context, purpose and/ or new service. We will explore EOSC in future in case new functionality is going to be implemented as part of the Gnomee web application.

Limitations and future improvement

OpenAIRE AAI service supports all main authentication providers and as such makes it easy for users to select the provider they wish to use. No limitations at this point.

Useful material related to this story

athenarc.gr

EOSC Future

Monitoring Tropical Forest Recovery Capacity Using RADAR Satellite Images

An EOSC in Practice Story where time series of satellite radar images are collected to measure forest recovery capacity at unprecedented spatial and temporal scales.

The Project Involved

C-SCALE

The C-SCALE project (funded from the European Union's Horizon 2020 research and innovation programme under grant agreement No 10101752) aims at enhancing the EOSC Portal with pan-European federated Copernicus data and computing infrastructure for their processing and analysis. The project provided resources which proved to be extremely useful for the RETURN use case. This aims at exploring time series of satellite radar (Sentinel-1) images from the EU Copernicus Earth Observation (EO) Programme to understand the recovery capacity of the Amazon rainforest. The outcomes of the RETURN research are important to help identify areas with slower forest recovery in the Amazon basin and potentially understand their causes.

The Context

Recent extreme droughts combined with accelerating human exploitation are pushing tropical forests to the point where they cannot recover, making them vulnerable to large unprecedented wildfires. This causes an urgent need to monitor the recovery capacity of tropical forests.

The Challenge

While time series-based break detection approaches have demonstrated potential to measure tropical forest recovery capacity, they have not yet been applied over large amounts of satellite data. The reasons for this are twofold:
- These methods require expert knowledge and are very compute intensive
- They cannot be easily scaled over large amounts of satellite data

The radar images are indeed collected every six days and have a pixel size of 10 m, resulting in terabytes of data when a period of several years is analysed over a large area such as the Amazon basin. Therefore, the main challenge for users is how to efficiently store, access, and analyse such big spatio-temporal data. There are different big satellite data processing platforms available, but many of them are not flexible enough, as they do not support user-defined codes or a complex workflow involving different software solutions.

Milutin Milenković

Research Scholar, International Institute for Applied Systems Analysis, and Associated Researcher, Wageningen University & Research

"The radar images are collected every six days and have a pixel size of 10 m, resulting in terabytes of data when a period of several years is analysed over a large area such as the Amazon basin. EOSC services help users face the main challenge to efficiently store, access, and analyse such big spatio-temporal data"

The solution

To address this challenge, the SURF computing infrastructure, i.e., the SPIDER cluster and the Sentinel-1 data cube prepared by the EODC GmbH was used. Those resources were available within the C-SCALE project as a part of a use case defined and designed together with EOSC experts. The interaction with the computing infrastructure and data was established through ssh connections and Jupyter Notebook. The RETURN use case researcher, Milutin Milenković has developed a Python code to analyse satellite radar image time series and used tools such as Jupyther and slurm to test and upscale the analysis on the cluster. Once the analysis is done, all the results and the code with usage examples will be available on the following GitHub repository

EOSC Future

The researcher

Milutin Milenković is a Reserch Scholar at the International Institute for Applied Systems Analysis, Novel Data Ecosystems for Sustainability group, and Associated Reseacher at Wageningen University & Research in the Geo-information Science and Remote Sensing Lab. His current research focuses on combining earth observation and citizen science for monitoring environmental change on the global scale.

Why do I need EOSC?

EOSC played a crucial role in Milutin's study by hosting big satellite data and providing computational capacity to analyse those data. Furthermore, EOSC cloud computing experts helped Milutin onboard his code to the EOSC computing infrastructure. The EOSC services helped overcome the challenge of processing and analysing terabytes of satellite images. As the results and code will be freely available, the whole computational setup and the outcomes will be fully reproducible, ensuring credible information about environmental change and nurturing open science practices. Moreover, the study will provide highly detailed and spatially-explicit information about signal disturbance and recovery over the Amazon basin, enabling studies on the relationship between areas with different signal recovery properties and their potential causes, i.e., contributing to the understanding of the resilience of the Amazon rainforest.
The EODC JupyterHub for global Copernicus data is available on the EOSC Marketplace

Across disciplines

The study's outcomes will be relevant for ecologists, foresters, and the remote sensing community. The computing approach will be relevant for the remote sensing community as an example of a custom analysis upscaling over a larger area. Furthermore, ecologists and foresters can further explore the study results, for example, by analysing spatial patterns and statistics of signal recovery to gain a different look at the forest disturbance and recovery process in the Amazon basin.

Useful material related to this story

wenmr.eu

Future developments

The study's results, i.e., spatially-explicit features of signal recovery, can be further explored and analysed by other experts. Those features can be an input for machine learning models to map and analyse causes of forest disturbance or predict follow-up land use classes. EOSC can foster such cross- disciplinary analysis by hosting the study results close to their computational resources and onboarding the code (analysis) from other researchers.

Useful material related to this story

- EODC JupyterHub for global Copernicus data

Empowering Global Collaboration in Structural Biology and Life Sciences

Unifying Research Teams and Streamlining Computational Approaches for Advanced Data Analysis and Modelling

Alexandre Bonvin

Scientific Director of the Bijvoet Centre for Biomolecular Research and Computational Structural professor at Utrecht University

"EOSC had a pivotal role in distributing the computational load by providing services that enabled job execution and computation. Furthermore, EOSC services facilitated user registration and authentication mechanisms, making it possible for researchers to access the tools provided by WeNMR."

The Project Involved

we-nmr

WeNMR is a Virtual Research Community which aims at bringing together complementary research teams in the structural biology and life science areas into a virtual research community at a worldwide level and providing them with a platform integrating and streamlining the computational approaches necessary for data analysis and modelling.

The Research Community

WeNMR serves a community of structural and computational biologists and life scientists distributed worldwide, covering over 135 countries. Researchers using WeNMR portals range from bachelor and master students through PhD students and post-doctoral researchers to advanced researchers. The majority of users are in academia and non-profit organisation, with some also coming from industry (pharma and biotech).

The Challenge

Structural biology is a field that focuses on understanding the intricate details of biological macromolecules at the atomic level, including their structural and dynamic properties. This knowledge is essential for understanding the majority of cellular processes and has various applications in fields like health and food sciences. The WeNMR initiative aims to provide valuable tools to the structural biology community while ensuring their accessibility in a user-friendly, efficient, and cost-effective manner. The challenge is to execute these tools in a distributed environment that optimizes their utilization, thus benefiting the wider scientific community.

EOSC service or tool used

Since 2009, WeNMR has collaborated with EGI, benefitting from a dedicated Competence Centre involving user experts in the scientific domain, resource centres, and the NGIs, the technology providers of the EGI, set up to further support the community needs. This activity allows the scientific communities to use High-Throughput Computing and Online Storage services to develop portals for life and brain scientists worldwide, also accessible on the EOSC Marketplace. EGI provided WeNMR with technical support to extend COVID-19-related research in the US using the computing resources provided by Open Science Grid. Additional consultancy and technical support were also offered to integrate the EGI Workload Manager in the WeNMR portals to facilitate the distribution of computing jobs.

Useful tips & tricks

EOSC services and tools have been implemented and refined over many years, in most cases in the context of European e-Infrastructure projects in which EGI was a partner or coordinator. Close collaboration with EGI experts over the years has been key to the smooth operation of WeNMR services. WeNMR services are described in the following publication: R. Vargas Honorato, P.I. Koukos, B. Jimenez-Garcia, A. Tsaregorodtsev, M. Verlato, A. Giachetti, A. Rosato and A.M.J.J. Bonvin. Structural biology in the clouds: The WeNMR-EOSC Ecosystem. Frontiers Mol. Biosci. 8, fmolb.2021.729513 (2021).

Benefits and impact

The HADDOCK service, one of the services provided by WeNMR, has attracted an increasing worldwide community of users with exponential growth since 2008. The impact is only expected to grow in the coming years. In a recent survey (2022) on the users who submitted jobs since 2020 (~7,500, ~660 responses collected), 27% of users indicated they used HADDOCK for education and 95% for research. This indicates that the training of next generation scientists in using modern e-Science EOSC solutions results in them using those later on in their research. The impact on research is also evident from the responses, with 56% indicating they used WeNMR tools in disease/ health-related research, 34% for drug and antibody design and 22% for COVID-19-related research.

Why do I need EOSC?

To achieve its goal, the WeNMR initiative developed a services suite comprising a front-end and a backend web page consisting of different software and scripts. EOSC had a pivotal role in distributing the computational load by providing services that enabled job execution and computation. Furthermore, EOSC services facilitated user registration and authentication mechanisms, making it possible for researchers to access the tools provided by WeNMR.

Across disciplines

WeNMR services are mostly used in disease/health-related research, but also find applications in food and material research, and have also become embedded in various educational programmes.

Useful material related to this story

wenmr.eu

Limitations and future improvements

WeNMR is working on making its services independent from the underlying hardware and operating system and easily deployable, for example, in EOSC cloud resources. This is an area where, with the support of EGI, WeNMR hopes to automate such deployments and scale up the services when needed.

Share your own #EOSCinPractice story here

EOSC Future

Implementing a seamless data transfer mechanism between different Research institutions

An EOSC in Practice Story where has been implemented a tool to access and share valuable data across borders and institutions.

The Project Involved

DICE

DICE (DATA INFRASTRUCTURE CAPACITY FOR EOSC) is a European Horizon 2020 project funded under Grant Agreement no 101017207. The project aims to develop and improve data management infrastructure for the European Open Science Cloud (EOSC). The project goals is to create an integrated and sustainable infrastructure that will enable researchers to access, share, and preserve research data more efficiently. DICE brings together a consortium of partners from across Europe to develop new tools, services, and policies for data management, and to work towards a more open and collaborative research environment.

CompBioMed, an EU-funded Centre of Excellence in Computational Biomedicine is focused on the use and development of computational methods for biomedical applications. The need for safe data replication and large data transfer between HPC centres, supporting a FAIR data cycle, and the strive to use existing tools for data management within the EU has resulted in a collaboration to set up a B2SAFE federation among HPC centres as core partners in CompBioMed.

The Challenge

The main challenge addressed by the federated B2SAFE service is the lack of a seamless data transfer mechanism between different institutions, which has hindered collaboration and research progress. The ability to share data sets across borders and institutions is essential for researchers to build on each other's work, ultimately leading to more impactful and innovative research. Besides the technical solution that we put in place, there are still challenges from legal and governance perspectives that need to be addressed for sharing data such as having risk assessments and data-sharing agreements between centres.

Nadia Tonello
Head of Data Management at BSC

James A J Wilson
Head of Research Data Services at UCL

Narges Zarrabi
Sr. Advisor & Product Manager at SURF

"The lack of a seamless data transfer mechanism between different institutions has hindered collaboration and research progress. B2SAFE service enables seamless data transfer between the institutions, making it easier for researchers to access and share valuable data across borders and institutions"

The Users

This EOSC in practice story targets three main types of users: (1) **citizens and citizen scientists** who gather significant data, (2) **researchers** who utilize and gain advantages from the data obtained through smartphones and sensors by citizens, and (3) s**oftware developers**.

The solution

In 2023, the federation of the B2SAFE service by the Barcelona Supercomputing Center (BSC), University College London (UCL), and SURF (the Dutch national e-infrastructure provider) represents a significant milestone in the research and innovation landscape. The seamless data transfer enabled by the federated B2SAFE service makes it easier for researchers to access and share valuable data across borders and institutions, facilitating collaboration and knowledge exchange. This achievement demonstrates the growing importance of international collaboration in research, as well as the critical role played by advanced data management services like B2SAFE in supporting research excellence and innovation. The federated B2SAFE service provides researchers with a powerful tool to overcome the barriers to data sharing and collaboration, ultimately accelerating the pace of scientific discovery and innovation.

The EOSC Future project is co-funded by the European Union Horizon Programme call INFRAEOSC-03-2020 - Grant Agreement Number 101017536

58

EOSC Future

The service providers

The Barcelona Supercomputing Center (BSC) is a research center that specializes in high-performance computing, data analytics, and artificial intelligence. It is located in Barcelona, Spain and collaborates with national and international partners on various scientific projects.
SURF, is the Dutch national infrastrcuture provider that support for supports research and education in the Netherlands. SURF is a service provides for the B2SAFE service and contributes in internation projects related to EOSC.
UCL is a university with over 16,000 staff and 50,000 students enaged in teaching and research across a broad range of academic disciplines. It is home to a significant proportion of the research being undertaken as part of the CompBioMed projects. It's institutional data management infrastructure is used to store and process some of the project's data, including sensitive data held on the UCL Data Safe Haven.

Why do I need EOSC?

EOSC has emerged as a game-changer in the research community, providing unparalleled visibility and accessibility to research projects across Europe. With an emphasis on interoperability, EOSC has opened up new avenues for collaboration and data sharing, making research more efficient and effective. B2SAFE service play a crucial role in enabling researchers to manage and share their data securely and efficiently, while also catering to the specific requirements of various specialized disciplines. CompBioMed is one of the communities that has participated in a pilot to setup the federated network of B2SAFE nodes, with an aim to help streamlining data management and enhance research outcomes.

The impact on society

The federated B2SAFE service is a powerful tool that not only benefits its target users but also has the potential to create a broader positive impact on society. The ability to facilitate seamless data transfer and collaboration between researchers from different institutions can accelerate the pace of scientific discovery and innovation, leading to the development of new technologies, treatments, and solutions that benefit society at large. The federated nature of the B2SAFE service further enhances its capabilities, allowing it to connect researchers across different regions and disciplines, promoting collaboration and sharing of knowledge. As a result, the B2SAFE service has the potential to play a critical role in addressing some of the most pressing challenges facing society today.

Useful material related to this story

DICE.eu

Across disciplines

In addition to supporting its target users, the federated B2SAFE service has significant potential in promoting cross-disciplinary research initiatives. By enabling researchers from different disciplines to collaborate and share their data sets seamlessly, B2SAFE can break down the silos between different fields, leading to the integration of knowledge and expertise from various domains. This cross-disciplinary collaboration can ultimately lead to the development of more innovative and impactful research, with the potential to address some of the most pressing challenges facing society today. The federated nature of B2SAFE further enhances its capabilities in promoting cross-disciplinary research, as it can connect researchers from different regions and institutions and foster a more collaborative research ecosystem.

Future developments

B2SAFE can be expanded with new features, like the federation with other repositories, the connection with other data management services like B2SHARE and B2FIND, to complete the lifecycle of the data.

Sustainability for an EOSC in practice

B2SAFE is meant to be used as large mid-long term storage for data exploitation. The cost of the service has been partially covered by the INFRA-DEV-07 DICE EC project, under Virtual Access, until the end of the project.

Future funding model scenarios

For BSC, the VA mechanism has been just the entry mechanism, but it is not sustainable, given the mid-long term nature of the data service. The service in Spain can be provided within the data management services of the RES (Spanish Supercomputing Network) free at the point of use for Spanish research groups.
SURF offers the B2SAFE and its federations as a service in production. The service is connected to the tape archive in the backend for long-term preservation of the stored data. The VA mechanism was used during the DICE project for granting access. After that requests for access can be made through computing and data services calls funded by the Dutch National Funding Organisation for Research (NWO) via this portal.
Other providers federated with our storage have their own specific financial plan.

Prototyping new innovative services

Stories from projects funded under call INFRAEOSC-02-2019

EOSC Future

Connecting researchers, developers and citizen scientists in a unique mobile app environment.

An EOSC in Practice Story where heterogeneous data are collected via mobiles or sensors, and made accessible in a secure, open and easy way.

The project involved

Cos4Cloud

"What we frequently experience with funded projects is their time limitation. Projects last mostly 4 years, then they end, the funding ends… and the community is gone. This is why we really started early in the project to think about its future sustainability"

Norbert Schmidt, Owner @DDQ & Partner @Cos4Cloud

Cos4Cloud **(Co-designing Citizen Observatories Services for the EOS-Cloud)** is a European Horizon 2020 project funded under Grant Agreement no 863463. The project aims at boosting citizen science technologies. One of the biggest challenges of citizen science is the **quality of data**, as well as maintaining the citizen observatories used to collect this data. Cos4Cloud is addressing these challenges by developing twelve technological services to improve citizen science platforms.

The Users

This EOSC in practice story targets three main types of users: (1) the **citizen and citizen scientists** who collect valuable data, (2) the **researchers** who use and benefit from the data collected by citizens via smartphones and sensors, and (3) **software developers**.

The Challenge

Currently there are **thousands of citizen science apps constituting a fragmented ecosystem**. Each app comes indeed with its own login system and with a separate database accessible to specific groups of researchers. Most of these apps are still listed in app stores, but are actually abandoned, since they were created only for the purpose of a project. Beyond all concerns coming from these apps being unsustainable, it is also frustrating and time consuming to use them in such a disaggregated manner: users need to login multiple times, face the complexity of having data stored in different places and with different formats. **An integration service is what could solve such problems.**
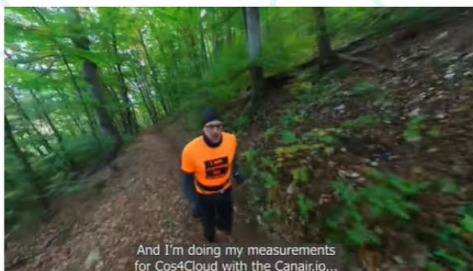
The solution

A MOBIS user measuring air quality near Brasov (Romania)

The proposed solution is called **MOBIS (Mobile Observation Integration Service)** and it provides a service that makes it easier to gather all kinds of data, and keep them open. MOBIS collects and processes data from different mobile observatories (most of them via smartphones with or without sensors). It can be used as a back end service for (citizen-) science apps with a single login. MOBIS is openly available to researchers & citizens. The final aim is for MOBIS to become a large repository for citizen science data, including environmental and bio diversity data or air and water quality measurement, to give a few examples. MOBIS can combine a multitude of heterogeneous data, offer interoperability with other apps and take care of GDPR compliance regarding the (re-)use of data and images. As of today, three citizen apps are already integrated with MOBIS.

The service provider

The MOBIS service is provided by DDQ: Pocket Science , a small Dutch company specialised in the development of mobile-based (citizen-) science applications with research partners worldwide. The company also offers smartphone back ends to facilitate mobile cloud computing, machine learning, user notifications, storage and offline synchronisation for working in remote areas. DDQ is an EOSC provider, whose services are hosted on the EGI High Performance Infrastructure. DDQ has developed MOBIS in the Cos4Cloud framework.

Why do I need EOSC?

MOBIS service brings the following benefits to its users thanks to its presence on the EOSC Portal Catalogue and Marketplace :

» unique mobile repository for research data accessible throughout Europe
» integration of heterogeneous data
» integration of multiple apps & easy onboarding process with minor devlopment skills required
» GDPR compliance

At the same time such a service provides the following benefits to the provider:

» New partnership opportunities
» Further visibility at European level

You can Access MOBIS on EOSC Portal Catalogue and Marketplace here

The impact on society

MOBIS' effect on society is to facilitate the interest and participation of **multiple users in science** in the short run. In the long run, the idea is to provide continued support for this service to still exist after the project ends, so that the involvement of users into science through the sharing of a complete open set of cross-disciplinary data can be sustained and grow over time. But there is more: accessing precise data about the environment can be used to monitor problems such as pollution in cities. Municipalities can access this information and take policy decisions to improve cities' liveability. This is just one example of the many impactful consequences of investing in a long-term open data storage and sharing service like MOBIS.

Across disciplines

This EOSC in practice story showcases cross-disciplinarity, given that the types of data that can be collected and shared via MOBIS encompass various fields, such as astronomy, environment and biodiversity.

Future developments

In its future versions, MOBIS will likely be improved with the addition of Machine Learning techniques, able to leverage on Artificial Intelligence to detect, for instance, the quality of water or other subjects by analysing its colour. More technological updates will possibly follow, including the use of quantum computing. Such developments will serve researchers, providing additional insights on the collected images and data, speeding up computing times and boosting efficiency via automation of some phases, such as image detection and recognition.

Sustainability for an EOSC in practice

Creating a sustainable service was one of the real triggers for MOBIS creation in the fragmented mobile app landscape. Cos4Cloud has equipped MOBIS with a sustainability plan being continuously updated which includes:

» the intention to have MOBIS hosted beyond the project's end. Possible options include keeping MOBIS on the EGI-ACE infrastructure or shifting MOBIS to the EOSC Exchange, provided that both data storage and adequate processing power are provided to support MOBIS service in both scenarios.
» The request of maintenance fees to apps that wish to stay onboarded on the MOBIS Framework. DDQ will carry out the fees through individual Service Level Agreements (SLAs). Some are already in place with citizen observatories.

Future funding model scenarios

Although a specific pricing for the maintenance of onboarded apps has not been defined yet, it is envisioned that the required fees will be asked to citizen science projects and citizen observatories and depend on the app requirements and functionalities. Some apps generate enormous datasets and require large storage and processing power, while other apps are doing most of the processing on the phone or are only data-entry apps, the fee will be of approximately 1.000€ per year for apps with basic requirements and increase with the complexity of requirements, such as dedicated back-end services, dashboarding, scripting, notification, connection to Artificial Intelligence or Machine Learning modules.

In addition, DDQ: Pocket Science is looking for and already partnering with other open data initiatives, including EOSC-related ones. The objective is to build a business roadmap for growth and sustainability by April 2022, which is focused on products and services instead of projects and assignments.

Useful material related to this story

» MOBIS: An EOSC service to create integrative environmental and biodiversity citizen science apps
» Run4Science, a set of MOBIS use cases where passionate sports persons perform in field data acquisition in raw natural spots
» MOBIS Framework

Want to learn more about the other services being developed by **Cos4Cloud**? Read here
Liked this **#EOSCinPractice** story? Follow @EOSCFuture for more!

EOSC Future

Supporting knowledge creation and sharing by building a standardised interconnected repository of biodiversity data.

An EOSC in Practice Story where a biodiversity data community is built according to FAIR principles.

The project involved

Cos4Cloud

[Cos4Cloud](#) **(Co-designing Citizen Observatories Services for the EOS-Cloud)** is a European Horizon 2020 project funded under Grant Agreement no 863463. The project aims at boosting citizen science technologies. One of the biggest challenges of citizen science is the **quality of data**, as well as maintaining the citizen observatories used to collect this data. Cos4Cloud is addressing these challenges by developing twelve technological services to improve citizen science platforms.

The Users

This EOSC in practice story targets a very wide user base as it is addressed to any researchers, teachers, students, companies, institutions and, more generally, anyone interested in knowing, studying or analysing biodiversity information.

The Challenge

Citizen Observatories are currently faced with **fragmentation problems**. The collected data is heterogeneous and comes in varied formats. This is an obstacle for the users who have to dedicate massive resources to elaborate, standardise and aggregate data, thus making the data collection and management phases very long and inefficient. These problems relate to the difficulty of practically implementing FAIR (Findable, Accessible, Interoperable, and Reusable) rules, which instead is a crucial feature for the integration of the services in the EOSC marketplace. To solve such challenge and support users when downloading and using the data, **a better dialogue between the different citizen observatories is needed**.

The solution

Cos4Bio is a **co-designed, interoperable and open-source** service that integrates biodiversity observations from multiple citizen observatories in one place, allowing experts to save time in the species

"Adopting co-design and co-design thinking approaches to develop Cos4Bio was a real challenge. However, involving developers, experts and users in the early phases of the work was fundamental to really help us build a better product"

Santiago Martinez de la Riva, CEO, @Bineo Consulting **& Cos4Bio Developer** @Cos4Cloud

View of the Cos4Bio service website
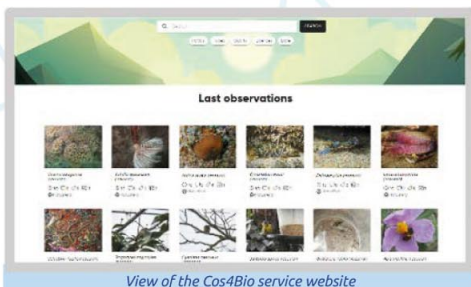
identification process and get access to an enormous number of biodiversity observations. Co-design principles were followed in the creation phase to ensure a collaborative and interdisciplinary approach that could maximise the service quality and usability. Cos4Bio allows citizen science experts to view and identify all observations from a single place, interacting with the community and contributing their knowledge about each species. When an observation has been identified in Cos4Bio, this information is updated in the citizen observatory where it was published, awaiting final validation according to the algorithm defined in each observatory. Cos4Bio has a search system that allows to consult the observations of a specific species or a specific location across different citizen observatories. It is also possible to apply criteria to filter information such as origin, type, quality of observation, license or date. All the information can be downloaded in "csv" format following Darwin Core (DwC), a widely known and accepted standard within the biodiversity community. Finally, Cos4Bio relies on Authenix to provide secure and federated authentication services to the users and GBIF Backbone Taxonomy to manage data classification.

The service provider

The Cos4Bio service complies with FAIR principles and is provided by Bineo Consulting, a small Spanish software company specialised in the offering of Web Development, Mobile, Infrastructure, Big Data and Data Science services. Bineo Consulting is an EOSC provider. Bineo Consulting leads the work on Interactive Services in Cos4Cloud.

Why do I need EOSC?

Cos4Bio service brings the following benefits to its users thanks to its presence on the EOSC Portal Catalogue and Marketplace :

» one unique repository for citizen collected data on biodiversity accessible throughout Europe
» integration of heterogeneous data following accepted standard formats
» integration of multiple nature apps & repositories
» provision of additional services via integration with other resources available on EOSC, such as Authenix and Pl@ntNet
» GDPR compliance
» fast identification of the species and access to data
» impact monitoring available for each entry about downloads, notifications and comments
» visibility to all contributing citizens and researchers that can share their profile and details about the data they have provided

At the same time, being the service accessible via EOSC, the following benefits are left to the provider:

» Increased number of observations
» Improved identification of species
» Larger data than a single observatory to maximise algorithm training potential
» Higher quality of information
» Increased community

You can Access Cos4Bio on the EOSC Portal Catalogue and Marketplace here

The impact on society

Cos4Bio's effect on society is to facilitate the interest, participation and connections of **multiple users and experts in biodiversity and environmental science**. The ultimate mission is to create a wide and engaged community facilitating knowledge creation and sharing.

Across disciplines

Biodiversity is not only flora and fauna. It has important interactions with various environmental data. In this sense, the information provided via Cos4Bio can be used as part of wider environmental studies.

Future developments

Bineo Consulting is developing in the Cos4Cloud framework two other services to follow-up **Cos4Bio**. Cos4Env and DUNS (Data Use Notification Service) are also being released and onboarded to EOSC. Cos4Env, in particular, will have a similar structure and purpose to Cos4Bio, despite being broader in scope. All types of environmental data (e.g.CO2 emissions, temperature, humidity) coming from citizen observatories will be included in this service and not only biodiversity. Finally, DUNS will act as a centralised service to (1) register usage of the citizen science observations downloaded from the Cos4Bio and Cos4Env portals and (2) make this information available to the citizen observatory the observation comes from. The aim is to help make citizen observatories aware of how their data is used and reward their users' contributions.

Sustainability for an EOSC in practice

Nature apps and services, such as Natusfera and Pl@ntNet are already integrated in Cos4Bio. Others, including ArtPortalen and iSpot, will be integrated soon to help enlarge the community. According to Santiago, creating a strong community is the first step to make this service scalable and sustainable in the future. In order to facilitate the building of a massive scale, Cos4Bio is working on multiple levels:

» Facilitating and speeding up the onboarding and integration of new Citizen Observatories via easy interoperability mechanisms, such as APIs
» Establishing connections with other services developed in CoS4Cloud and available on the EOSC Portal Catalogue and Marketplace, so to create an ecosystem of biodiversity services
» Working on communication, dissemination and engagement of users and experts via recognition mechanisms and incentives to the use of the service. These include visibility for the experts, dashboards, reporting and other tracking systems to monitor the performance of each contributor

Future funding model scenarios

Cos4Bio wants to invest as much as possible on growing the community and quality of the content and data provided, connecting Cos4Bio to a wider ecosystem of services, also currently under development, such as the ones mentioned in this EOSC in practice story. Such actions should be undertaken during the life of Cos4Cloud project, so that one of the possible funding scenarios is the natural support of this unique data source via a follow-up project or similar EU-financed grants.

An alternative or concurrent funding scenario includes a yearly fee-based mechanisms for the Citizen Observatories that want to integrate their observations in Cos4Bio, while the service shall remain free for the experts and researchers using the data available on Cos4Bio. The fee paid by the Citizen Observatories would be inversely proportional to the number of observations provided.

The sustainability plan and business model of this and the other services produced under the remit of Cos4Cloud will be finalised in July 2022.

Useful material related to this story

» Cos4Bio website
» Cos4Bio Q&A

Want to learn more about the other services being developed by **Cos4Cloud**? Read here

Liked this **#EOSCinPractice** story? Follow @EOSCFuture for more!

EOSC Future

Supporting cross-disciplinary research in natural sciences

An EOSC in Practice Story on Pl@ntNET, a Citizen Observatory and AI-based plant identification solution.

The project involved

Cos4Cloud

Cos4Cloud (Co-designing Citizen Observatories Services for the EOS-Cloud) is a European Horizon 2020 project funded under Grant Agreement no 863463. The project aims at boosting citizen science technologies. In particular, it addresses one of the biggest challenges in citizen science: the **quality of data**, as well as maintaining the citizen observatories used to collect this data. Cos4Cloud is tackling these challenges by developing thirteen technological services.

The Challenge

This story derives from a market need identified by the service provider. There are thousands of applications in the natural sciences field who were asking for an automated plant identification system within their own applications. Examples include apps that need plant identification systems because they use this information to study the properties of soil and its quality, or because they need to identify specific agricultural practices or promote biodiversity. There are different types of services that build on plant identification which is therefore horizontal to the needs of many related communities. The developers of such apps need to access plant information easily to include this in their research or commercial workflow. The challenge of providing this service mainly lies in the very high diversity of plant species to be identified and in managing the large number of queries and connected users without destabilising the system.

The solution

The proposed solution is called Pl@ntNET, a citizen observatory and AI-based Platform designed to monitor plants biodiversity and help identify plants via pictures. It is organised in different thematic and geographical floras operational via a search engine.

One of the tools developed from this service is a mobile app for final users that allows browsing plant species, seeing images and voting people's observations. In addition, Pl@ntNET has developed a visual
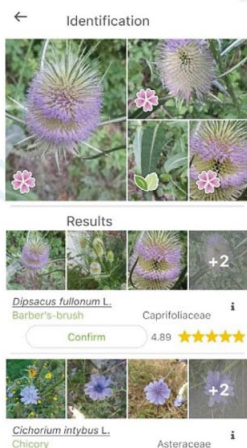
Pl@ntNET mobile app overview

identification engine as a web service in the Cos4Cloud project framework. This web service allows citizen observatories, startups or companies to integrate Pl@ntNet's visual identification tool into their own solutions by using an API. The identification service is updated monthly based on new data produced, shared and validated by the network of registered participants. Pl@ntNet mainly concerns wild plants that have propagated spontaneously in the natural environment, but also cultivated plants of agronomic and horticultural interest. The visibility and use of this platform has accelerated since February 2013, after deployment on mobile devices. Since 2013, the number of daily users has doubled every year, reaching more than 400,000 users per day at some peaks in 2021.

The users

The user community of Pl@ntNet consists of

(1) the **citizen and citizen scientists** who are interested in identifying and knowing more about plants, and

(2) **developers** of applications and software for plant identification, gardening, biodiversity management, agroecology or other related fields.

The service provider

The solution is provided by a consortium of French research organisms piloted by Inria, the French national research institute for digital science and technology. As a technological institute, Inria supports the diversity of innovation pathways: from open source software publishing to the creation of technological startups. With its regional centres, Inria has a presence at leading research university campuses, throughout the regions of France and the country's industrial and entrepreneurial ecosystems for digital technology.

Why do I need EOSC?

Pl@ntNet identification service already existed before its publication on the EOSC Portal Catalogue and Marketplace. However, providing Pl@ntNet via EOSC has brought these benefits:

» Further visibility and credibility at European level
» Possibility to reach a new audience, i.e. European research projects and scientific community

You can Access Pl@ntNet identification service on EOSC Portal Catalogue and Marketplace here.

The impact on society

Pl@ntNet's effect on society is to facilitate the interest and participation of multiple users in science and support dissemination of results and knowledge sharing in the fields of environmental education, agroecology and gardening.

Across disciplines

The types of data that can be collected and shared via Pl@ntNet encompass various natural science fields, such as agriculture, environment and biodiversity that were in need of a plant identification service to complement their research workflows and tools.

Future developments

In its future versions, Pl@ntNet will include the possibility to identify plants from pictures where more species are depicted together, one near the other. At the moment instead the service only recognises one plant at a time. In this way it will be possible to let the service identify plants that are photographed by robots or drones, important for precision agriculture.

Sustainability for an EOSC in practice

Creating a sustainable service is one of the requirements for EOSC future development. In this specific case, the service was already created and sustained by the service provider through the following main funding sources:

» Donations are collected from the users of the service. Last year, more than €200,000 were collected via donors that spontaneously wish to support the service which is free to use. Such donations are solicited via promotional campaigns
» The research consortium hosting Pl@ntNet directly supports it financially
» Participation in European research projects is the third financial source.

The idea is to keep the current funding model. An additional lever for the overall sustainability will be given by the integration of this service with the other services that have been developed under the Cos4Cloud project.

Useful material related to this story

» Video description of P@ntNet
» Pl@ntNET FAQ list

Want to learn more about the other services being developed by **Cos4Cloud**? Read here

Liked this **#EOSCinPractice** story? Follow @EOSCFuture for more!

Support to the EOSC Governance

Stories from projects funded under call INFRAEOSC-05-2018-2019

EOSC Future

Supporting data findability, reliability and replicability in Earth, Meteorology and Climate science domains.

An EOSC in Practice Story on daily dust observations and forecasts from Europe, North Africa and the Middle East.

The project involved

EOSC SYNERGY

The dust observation and forecast services presented in this story are offered within the context of EOSC-synergy, which is a European Horizon 2020 project funded under Grant Agreement no 857647. The goal of the project is to integrate a set of multi-disciplinary thematic services into the EOSC ecosystem. The biggest impact brought by the EOSC-synergy project will be represented by the large increase in the number of resources, services and data repositories offered through EOSC.

> *"We need to go more in the direction indicated by the World Meteorological Organisation: implement a forecast portal where services like data visualisation, data analysis and data storage are provided."*
>
> **Francesco Benincasa**, Barcelona Dust Regional Centre Engineer @EOSC_synergy

The Users

This EOSC in practice story targets three main types of users: (1) **researchers using the uploaded data models** for their own research, (2) **national and international meteorological agencies**, disseminating forecasts to their users, (3) **policymakers**, providing official data to governments and international organisations to support the decision made by the governments.

The Challenge

The challenge addressed was to provide more reliable, findable and replicable data through dust related forecasts and observations to researchers and policy makers. Currently these users struggle to find good data sources, as they need to deal with multiple copies of the same records, records with many missing variables, compliance issues.
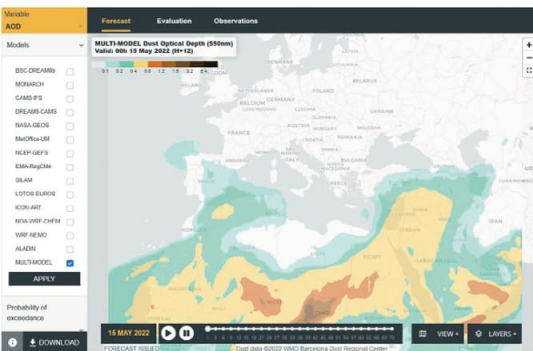
The solution

The dust observations and forecasts services are provided by the WMO Barcelona Dust Regional Centre. The centre was created in 2007 by the formal agreement of two Spanish institutions: the Meteorological State Agency of Spain (AEMET) and the Barcelona Supercomputing Center (BSC). The new services EOSC-based are still in their early stages, but the number of users, generic, in research, and public policy fields

is already increasing. In particular, the WMO Barcelona Dust Regional Centre, through its services, allows to simulate, thanks to numerical models, the forecast of dust and sand storms in Europe, North Africa and Middle East. Some numerical models are run in-house; in addition, outputs of models from institutions like NASA are collected. Currently, 12 data models are made available and more are foreseen in the near future. All these data follow international standards and formats to make them reliable and usable by an international audience. Users can choose between the following products:

» **Daily Dust Forecast**: this section includes daily forecasts, their comparison with observations, and dust-related observations. Data are displayed through interactive dashboards with three panels: Forecast, Evaluation and Observations. The dashboards allow users to centralize dust information from different sources (satellites, ground-based networks, experimental campaigns, marine ecosystems and models). All the services provided by the WMO Barcelona Dust Regional Centre are using EOSC related infrastructure, in particular B2SAFE and B2HANDLE, provided by EUDAT. To access storage services here, providers need to deliver the Persistent Identifiers (PID) of their data. This allows users to easily find files, web pages and any other resources they need.

» **Data Download Service**: This service provides almost all models outputs participating in the ensemble with two policies. Public access (2-days delay data) for all registered users and Restricted access (real time data) for partners.

» **Dust Products Catalogue**: Users can find centralized dust information from different sources (satellites, ground-based networks, experimental campaigns, marine ecosystems and models), that can be filtered by parameters, type of measurement or instrument and region.

Moreover, a climate service dashboard is under development and will be added soon.

The Daily Dust Forecast from Barcelona Dust Regional Centre

The service provider

Barcelona Supercomputing Center (BSC) is the national supercomputing centre in Spain, specialised in high performance computing (HPE). The centre, with a total staff of more than 725 R&D experts, is a point of reference for scientific communities that need HPC. Their research fields are: Computer Sciences, Life Sciences, Earth Sciences and Computer Applications in Science and Engineering.

Why do I need EOSC?

The Barcelona Dust Regional Centre services, called Daily Dust Products, bring the following benefits to their users thanks to their presence in the EOSC Portal:

» unique storage service capacity, with unique features in terms redundancy, federability and Persistent Identifiers generation.
» additional visibility to the service provided.
» opportunity to have a direct channel toward EU research audience
The Barcelona Dust Regional Centre services are already available on EOSC Marketplace! Access them here.

The impact on society

The Barcelona Dust Regional Centre services impact on society in different fields. In the health sector, for example, there are studies investigating the relationship between the frequency of dust storms in a given country and the likelihood to experience pandemics and illnesses like meningitis, as well as allergies and diseases of the respiratory system. There are also studies showing how dust in agriculture is used as fertiliser. In addition, dust is the most prominent component of atmospheric aerosols after sea salt. It impacts climate directly by interacting with both solar and terrestrial radiation, thus influencing the global radiation balance. By acting as cloud condensation nuclei, dust also impacts climate indirectly altering cloud microphysical structure and rainfall process. Understanding better all these relationships can open paths to new green sustainable policies.

Across disciplines

This EOSC in practice story showcases a service that provides data specific to the climate research domain. Nonetheless, connections with health, agriculture, visibility, aviation and transport domains are evident due to the impact dust concentration has on these areas. This could lead to multidisciplinary collaboration.

Future developments

In its future versions, all the Barcelona Dust Regional Centre services will likely be improved with the addition of new visualisation, forecasts and time series tools. Users would be able to use maps, with a time series feature, for producing accurate forecasts. A new functionality will be the implementation of a cross section with the visualisation features capable of considering the impact of the altitude. This additional data will provide new insights to researchers. In the prospective of continuously keeping on improving the services, a feedback system will be implemented.

Sustainability for an EOSC in practice

The Barcelona Dust Regional Centre services do not depend on EOSC financially. Users can access the services for free, they will only need to sign an acknowledgment of data use for research purposes. On the operative side, this service uses the EUDAT infrastructure, in particular B2SAFE and B2HANDLE and the storage capacity, from the current 50 TB, will probably be increased in the future.

Future funding model scenarios

Currently the Barcelona Dust Regional Centre services are funded mainly by the national funding coming from the Spanish national meteorological agency, AEMET. Since 2010 the service contract is renewed every 2-3 years. The products provided by WMO Barcelona Dust Regional Centre are well used in the European research community. Institutions based in UK, Finland and NASA are uploading the data produced by their models on the Barcelona Dust Regional Centre.

Useful material related to this story

» Barcelona Dust Regional Center User Guide
» WMO SDS-WAS Regional Center for Northern Africa, Middle East and Europe
» Global Assessment of Sand and Dust Storms by UNEP
» The paper in preprint on dust forecast reanalysis (climate application) for the DustClim project

Want to learn more about the other services being developed by **EOSC-synergy**? Read here
Liked this **#EOSCinPractice** story?
Follow @EOSCFuture for more!

EOSC Future

Supporting the implementation and adoption of EOSC by helping resource providers verify legal and ethics onboarding requirements.

An EOSC in Practice Story where a free tool is offered to prospective providers to self-assess adherence to EOSC rules of participation.

The project involved

NI4OS-Europe **(National Initiatives for Open Science in Europe)** is a European Horizon 2020 project funded under Grant Agreement no 857645. The project aims to be a core contributor to the European Open Science Cloud (EOSC) service portfolio, commit to EOSC governance and ensure inclusiveness on the European level for enabling global Open Science. It focuses on supporting the development and inclusion of the national Open Science Cloud initiatives in 15 Member States and Associated Countries in the EOSC governance.

"Since the EOSC Rules of Participation are aligned with FAIR principles, RoLECT is a very useful roadmap beyond the purpose of onboarding services to EOSC. It can indeed serve all researchers for all those issues related with fairness of resources and data in general"

Marianna Katrakazi, Legal Associate @ATHENA RC; Partner @ NI4OS-Europe

The Users

This EOSC in practice story is mainly addressed to **EOSC resource providers** seeking to verify whether the resources they wish to onboard (service or other) align with the EOSC rules of participation (RoP), focusing on the Legal and Ethics requirements. Secondly, it also helps **researchers and research organisations** in raising awareness on the various aspects of the RoP a resource must comply with, if aiming for EOSC onboarding.

RoLECT service access page

The Challenge

The current version of the EOSC rules of participation report edited by the EOSC Executive Board Rules of Participation (RoP) Working Group (WG) states the standards and conduct required of EOSC participants. The rules themselves are high level in order to be generally applicable and long-lived. However, the lack of details and specific compliance criteria can make it difficult for providers to understand all the needed requirements, if they have no legal background or experience, which is not unusual in many companies, especially SMEs. Moreover, openness, transparency, and inclusiveness are among the key principles that have been defined by RoP. Ensuring these principles can be a particularly demanding task when dealing with legal and ethics aspects. Therefore, **a self-assessment tool, which can help EOSC service providers understand the main priorities of EOSC RoP and the actions they have to undertake to be compliant with the core principles of the RoP, is needed**.

The solution

The solution proposed to adress the mentioned challenge is a tool called **RoLECT**.

It is designed to promote compliance and identify non-compliance with EOSC RoP and provides a set of questions that are categorized into three levels of importance (high, medium and low) marked with different symbols.

The questions have been classified according to the priorities set out in EOSC RoP. The symbols serve as indicators of compliance and help resource providers assess what information or omission of information is most important for verifying adherence with RoP.

Questions of **high importance** are critical as they reflect the basic principles of RoP.

Questions of **medium importance** regard information that is consistent with EOSC RoP, while **low importance** questions tackle information with a relatively low impact on the assessment result. If, after the assessment of the tool, the minimum standards and requirements are not met, the provider can verify what issues need to be addressed in order to align with the core principles of EOSC RoP. In addition, the assessment raises awareness about compliance criteria since it helps providers to identify any omissions, i.e any requirements of EOSC RoP included in RoLECT of which they may be unaware.

The service provider

RoLECT was developed by Athena Research & Innovation Center, whose aim is to build knowledge and devise solutions and technologies for the digital society. The company is one of the **NI4OS** partners and a member of the EOSC Association, which brings together key stakeholders in the European research environment to agree on strategies for the advancement of Open Science and to optimise the conditions for research outcomes, and ultimately, to make the European Open Science Cloud (EOSC) a reality.

Why become an EOSC provider?

Becoming an EOSC provider allows sharing resources beyond community boundaries. EOSC is indeed intended to make access to scientific data and other scientific outputs easier and more efficient by federating existing scientific data infrastructures and digital infrastructures for data exploitation that are now spread across disciplines and EU member states.

Providing services and tools to the EOSC portal, allows to:

» Advertise them on the EOSC Portal and promote their adoption outside traditional user groups, reaching a wider user base
» Get statistics about access requests and customer feedback
» Get a free online platform where you can manage service requests, interact with users and provide support to them, and agree the most suitable service levels.
» Allow users to authenticate with their own credentials to access your services and resources thus facilitating their adoption
» Contribute to the definition and maintenance of EOSC service provisioning policies and the portfolio roadmap.
» Join the group of providers that meet EOSC quality standards

RoLECT facilitates providers in establishing a presence on the EOSC Portal Catalogue and Marketplace, as it allows to:

» save time to make an efficient onboarding process
» have a useful roadmap for all issues related with fairness of resources and data, given that RoP are aligned with fair principles

Read more about why to become an EOSC Provider here.

RoLECT is available on NI4OS-Europe service catalogue, a regional catalogue through which all the project's services are onboarded to EOSC.

Access the RoLECT assessment tool as guest or authenticated user here.

What the users of the service say

 RoLECT is very useful since there are no similar tools that can assist resource providers before the onboarding process" (Librarians).

"RoLECT has been regarded as a useful input on legal matters; it can be leveraged to assess legal aspects related to different resource types" (TF charter Rules of Participation Compliance Monitoring).

The impact on society

The tool helps enriching the EOSC ecosystem with new resources, because it encourages resource providers to overcome legal and ethics barriers. It helps also to build trust in the resources available via EOSC and it raises awareness and legal and ethics requirements for FAIR research.

Across disciplines

RoLECT has a guided sequence of questions that are standardised and simplify the existing RoP. Those questions have been formulated in such a way that they are general enough to cover all resources or thematic topics regardless of the discipline. So, it helps provide clear legal and ethics requirements that foster cross disciplinary research, avoiding all those fragmented regulations that may apply only to some specific specific topics.

Future developments

The development team follows closely any discussion and progress on EOSC RoP. The assessment tool will be enriched based on the outputs of EOSC RoP Compliance Monitoring Task Force.

Sustainability for an EOSC in practice

NI4OS-Europe's aspiration is that this resource and the others being developed by the project will be part of a helpful toolkit for all EOSC providers even after the project's end. One important aspect related to making the resources sustainable is the fact that NI4OS-Europe offers all its tools as open source solutions available through GitHub, so that they can be taken up by the communities that are interested in using and further developing the tools

Future funding model scenarios

RoLECT directly addresses one of the major open issues for the EOSC Association, thus for the sustainability of the tool the discussions and outcomes of the EOSC RoP Compliance Monitoring Task Force are monitored closely. The detailed RoLECT sustainability and funding model will be released in the beginning of 2023.

Useful material related to this story

» RoLECT presentation at EOSC Symposium 2021
» (Re)watch the EOSC provider days
» Open Science Fair Demo

Want to learn more about the other services being developed by **NI4OS-Europe**? Read here
Liked this **#EOSCinPractice** story?
Follow @EOSCFuture for more!

Enabling the maximal re-use of data while ensuring IPR compliance.

An EOSC in Practice Story on the freely available NI4OS Licence Clearance tool, that facilitates clearing copyrights of any re-used research content.

The project involved

NI4OS-Europe **(National Initiatives for Open Science in Europe)** is a European Horizon 2020 project funded under Grant Agreement no 857645. The project aims to be a core contributor to the European Open Science Cloud (EOSC) service portfolio, commit to EOSC governance and ensure inclusiveness on the European level for enabling global Open Science. It focuses on supporting the development and inclusion of the national Open Science Cloud initiatives in 15 Member States and Associated Countries in the EOSC governance.
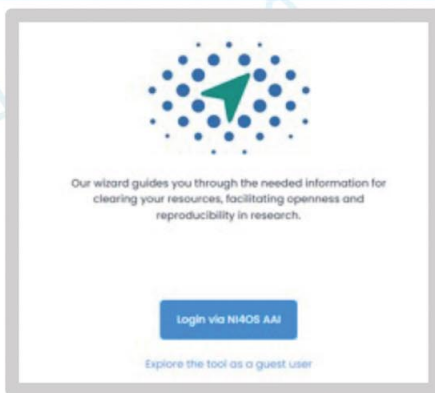
The Users

This EOSC in practice story is mainly addressed at **researchers or research organisations** who want to make their work available to the community, while protecting it with a compatible license.

The Challenge

Based on NI4OS-Europe's research, legal uncertainty about usage rights and high costs in clearance of licensing issues are very often a matter of concern when it comes to newly produced content but mostly the creation of derivative works, i.e. work that is produced based on other sources. Before sharing one's research it is important to correctly clear copyrights for any material that may have been used.

The solution

The **Licence Clearance Tool (LCT)** automates the licence clearance of derivative works and the selection of the most suitable licence for the users' resources. The tool provides a guided approach for establishing the proper open-source license required for the creation of a new (or synthetic) dataset, media, software etc. or for the re-use of existing unlicensed content. It proposes two workflows, one resource-driven for users that have a resource and want to clear derivate work licenses; the other is license-driven and it can be used when users have a target license and need to understand license compatibility with other resources.

"This tool encourages open science and FAIR research, while respecting IPR and supporting greater transparency, research reproducibility, dissemination, as well as the transfer of new knowledge"

Panagiota Koltsida, Senior Software Engineer, ATHENA RC

LCT service access page

The service provider

LTC was developed by Athena Research & Innovation Center, whose aim is to build knowledge and devise solutions and technologies for the digital society. The company is one of the **NI4OS-Europe** partners and a member of the EOSC Association, which brings together key stakeholders in the European research environment to agree on strategies for the advancement of Open Science and to optimise the conditions for research outcomes, and ultimately, to make the European Open Science Cloud (EOSC) a reality.

Why do I need EOSC?

Providing this tool on the EOSC portal, allows the tool to:

» Gain visibility at European level
» Reach a wider audience
» Reduce barriers to research sharing and re-use

The users benefit from an easy to use tool and do not need previous knowledge or a legal background to access the resource. Moreover the following benefits can be gained:

» Automation of the process of licence clearance
» Usability for different types of products (e.g. data, publications, software)
» No need for a legal background

LCT is available on NI4OS- Europe service catalogue, a regional catalogue through which all the project's services are onboarded to EOSC

Access the **LCT** wizard tool as guest or authenticated user here, and is expected to be onboarded to the EOSC Marketplace

What the users of the service say

"*We were very happy coming across the LCT in one of the OpenAIRE webinars, as it helps us to advise our postgraduate and doctoral students in assigning the proper license to the resources they upload in our institutional repository*". Librarian, Patras University Library

The impact on society

In line with the other services produced by NI4OS-Europe, LCT encourages open science and FAIR research, while respecting IPR. This translates into greater transparency, research reproducibility, dissemination, as well as the transfer of new knowledge.

Across disciplines

LCT can be used in different domains and helps deal with IPR issues, checking for compatibility among different licences. Its use can therefore support and facilitate collaboration among researchers in various fields.

Future developments

The current version of LCT provides guidance and supports 73 existing standard open-source licenses. The work, however, is continuously evolving and new functionality will be added including user login/ profiles and custom licenses in a future release.

Sustainability for an EOSC in practice

NI4OS-Europe's aspiration is that this resource and the others being developed by the project will be available even after the project's end. One important aspect related to making the resources sustainable is the fact that NI4OS-Europe offers all its tools as open-source solutions available through GitHub, so that they can be taken up by the communities that are interested in using and further developing the tools. Integrated in the NI4OS-Europe Catalogue, LCT can be used for clearance of licensing related issues during the onboarding process.

Future funding model scenarios

LCT is offered as a stand-alone tool, but also as part of the NI4OS-Europe Catalogue and onboarding processes, a fact that provides multiple sustainability pathways. The sustainability plans will be aligned with the current phase of the EOSC Association governance and the developments related to the EOSC Marketplace. The detailed LCT funding model will be released in the beginning of 2023.
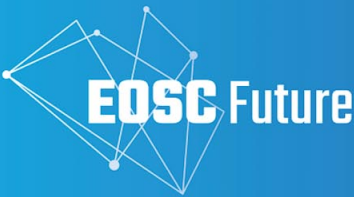
Useful material related to this story

» LCT: a license clearance tool for research outputs re-use
» Open Science Fair Demo
» LCT Flyer

Want to learn more about the other services being developed by **NI4OS-Europe**? Read here
Liked this **#EOSCinPractice** story? Follow @EOSCFuture for more!

EOSC Future

Supporting data trustworthiness using a repository policy generation tool

An EOSC in Practice Story of the Repository Policy Generator for data curators to manage repositories.

The project involved

NI4OS-Europe (**National Initiatives for Open Science in Europe)** is a European Horizon 2020 project funded under Grant Agreement no 857645. The project aims to be a core contributor to the European Open Science Cloud (EOSC) service portfolio, commit to EOSC governance and ensure inclusiveness on the European level for enabling global Open Science. It focuses on supporting the development and inclusion of the national Open Science Cloud initiatives in 15 Member States and Associated Countries in the EOSC governance.

The Challenge

While collecting data for the NI4OS-Europe landscape survey, it was possible to observe that a significant number of repositories from the partner countries were not registered with OpenDOAR and other registries and aggregators. The number of those lacking clear repository policies, not providing interoperability or addressing the policies incorrectly, was even greater. To be considered trustworthy, a repository should have a transparent policy, informing users about the roles, responsibilities, rights and procedures aimed at ensuring that deposited data are preserved and disseminated in line with the FAIR principles. Policies are also required in the process of onboarding platforms and services into NI4OS-Europe, OpenAIRE and EOSC.

The solution

The **Repository Policy Generator (RePol)** is an open-source web application developed by the University of Belgrade that facilitates the process of drafting repository policies. In **RePol**, policy documents are generated by filling out a web form: inputs and choices made by users trigger predefined sections of text and clauses to be added to the output document, which can be downloaded in a machine-readable format and edited. RePol'ssource code is available on GitHub. The tool can be used also to create not only repository policies but also more general policies and privacy policies.

"In order to create this tool we have collected the best practice policies for repositories. This is a step by step wizard that allows to easily create a policy for your repository, increasing its trustworthiness"

Eleni Toli, NI4OS-Europe Project Director, ATHENA RC

The Users

This EOSC in practice story is mainly addressed to **repository managers, librarians or researchers, especially at smaller research institution**s, that lack information on policies for repository access and management. Having correct policies is important to guarantee the trustworthiness of repositories.

The service provider

RePol – **Repository Policy Generator** was developed by the University of Belgrade Computer Centre (RCUB). RCUB is a computer centre whose primary function is to provide computer and information-communication services to faculties and other institutions members of the University of Belgrade. Besides that, RCUB represents a central communication node of the Academic network of Serbia (AMRES). The University of Belgrade is one of NI4OS-Europe's partners.

Branko Marović, NI4OS-Europe WP4 Leader, University of Belgrade

Why do I need EOSC?

Providing this tool on the EOSC portal allows to:

» Gain visibility at the European level
» Reach a wider audience
» Collect more feedback

The users benefit from an easy-to-use tool and do not need previous knowledge or legal background to access the resource

RePol is available on NI4OS- Europe service catalogue, a regional catalogue through which all the project's services are onboarded to EOSC

What the users of the service say

"We used RePol to update the policies of our institutional repository. It is a very easy-to-use tool that guides you through the whole process. We were happy to see that it didn't require legal knowledge to address even privacy-related policies" Researcher, University of Cyprus

The impact on society

While not directly linked to society as a whole, the purposes of this tool are in line with the open access movement. The policy generator facilitates access to information that maybe would not be possible or would not be so easy to find and access, and in this sense promotes open science principles. It helps providers in regulating the important aspects of service management. The publication of policies increases the transparency of the services and the users' awareness of openness and privacy protection concerns thus contributing to the development of a service-based economy.

Across disciplines

Cross-disciplinarity is directly addressed in this story because the tool offers policy options that are not domain specific nor attached to a specific discipline and can be used in many different fields or applied to cross-disciplinary repositories.

Future developments

As for the other NI4OS tools, **RePol** will be continuously updated and improved also thanks to its open-source code.

Sustainability for an EOSC in practice

NI4Os-Europe's aspiration is that this resource and the others being developed by the project will be available after the project's end. Key to the sustainability of RePol is that NI4OS-Europe offers all its tools as open source solutions available through GitHub so that they can be taken up by the communities that are interested in using and further developing the tools.

Future funding model scenarios

The future sustainability plan including possible funding model scenarios and actions will be developed in the coming months as the project approaches its end and will be released at the beginning of 2023.

Useful material related to this story

» Presenting RePol
» RePol Source code

Want to learn more about the other services being developed by **NI4OS-Europe**? Read here
Liked this **#EOSCinPractice** story? Follow @EOSCFuture for more!

Connecting ESFRI infrastructures through Cluster projects

Stories from projects funded under call INFRAEOSC-04-2018

EOSC Future

Developing an interoperable central hub for Social Sciences and Humanities resources.

An EOSC in Practice Story on the SSH Open Marketplace.

The project involved

SSHOC **(Social Sciences & Humanities Open Cloud)** is a project funded by the EU framework programme Horizon 2020 under Grant Agreement no. 823782, running from January 2019 to April 2022. It united **53 organisations** in developing the social sciences and humanities area of the European Open Science Cloud (EOSC).

The Users

The users in this EOSC in Practice story are **Social Science and Humanities (SSH) researchers** and **data experts** that need to find relevant resources and digital methods used in their specific research fields. In addition, the service described here can be also useful for **research institutions** and **policy makers** to have an overview of the tooling, service and resource landscapes in social sciences and humanities.

The Challenge

The Social Sciences and Humanities communities needed a central point to **gather and exchange information** about their tools, services, and datasets. Although plenty of project websites, service registries, and data repositories existed, they were mostly fragmented. The lack of contextualisation connecting these assets and offering domain-relevant means to discuss and enrich them was evident. Also, there was a need to **digitalise** this research domain which, in most cases, was not yet digital native. This siloed environment was the challenge SSHOC set out to bridge in January 2019.

The solution

The SSH Open Marketplace, **delivered in 2022**, is a discovery platform supporting researchers in the multi-faceted domain of humanities and social sciences at every step of the research data lifecycle. By grouping and connecting a wide set of tools, services, databases, training materials, workflows and publications on social sciences and humanities, the marketplace helps SSH academics and researchers to address their research questions providing these resources as "contextualised solutions".

> "Through the SSH Open Marketplace we are opening a bit the black box of digital methods to a wide range of research communities and the society in general, overcoming the fragmentation of resources characterising the social sciences and humanities"

Laure Barbot, DARIAH Project Officer

Preview of SSHOC Open Marketplace Service Catalogue

In addition, the marketplace works also as a community-gathering platform where all SSH researchers can contribute to the content, exchange information and feed on their experiences.

The service provider

The **SSH Open Marketplace** was developed in the SSHOC project. In particular, DARIAH ERIC coordinated the work for the creation of the marketplace, CLARIN ERIC took care of the interoperability aspects, and SSHOC Coordinator CESSDA ERIC supported the onboarding on the EOSC Portal. These three ERICs are maintaining the service relying on two main service providers: the Austrian Centre for Digital Humanities and Cultural Heritage (ACDH-CH) of the Austrian Academy of Sciences and the Poznan Supercomputing and Networking Center (PSNC) affiliated to the Institute of Bioorganic Chemistry of the Polish Academy of Sciences.

The development of this marketplace is a true collaborative and community effort where also the users have a significant role. By contributing to the tools and various resources offered in the portal, the users play a major role in ensuring the quality and continuous curation of this discovery portal.

Why do I need EOSC?

Providing this tool on the EOSC portal, allows to:

» reach a wider audience
» help researchers, as community contributors to the SSH Open Marketplace, make their research more visible across Europe
» collect feedback on the offered service
» upgrade the service level offered to users thanks to the services available for EOSC Providers

Moreover, the SSH Open Marketplace harvests SSH resources from the EOSC portal to provide SSH researchers with easy access to all available tools in their domain. That way, the users of the SSH Open Marketplace benefit from an easy to use hub and from updated information on how to use new resources, e.g. via the available workflows, which help researchers learn about new tools and save time to perform their research tasks.

The **SSH Open Marketplace** is available on the EOSC Portal , explore this resource!

What the users of the service say

"The SSH Open Marketplace is useful for the whole Social Sciences and Humanities community! The easier it is to browse information about existing resources, the better it is for everyone. It saves time and enables research. Curation of metadata in discovery portals matter and past experience has shown that it is difficult! The collaborative approach of the SSH Open Marketplace where anyone can contribute in enriching the records while browsing the catalogue and where a team of moderators supports its continuous curation, is an interesting way to overcome some of these challenges."

Amelia Sanz, María Robles and Adrián Menéndez de la Cuesta from Complutense University of Madrid.

The impact on society

The **SSH Open Marketplace** helps enriching the EOSC ecosystem with new resources from the social sciences and humanities domain. Being an open access platform, it allows anyone to browse, discover or contribute to the resources thus helping to open the "black box" of digital methods to the wider public. It provides insights that can be useful for a wide array of stakeholders including companies and SMEs who are working to provide services for researchers. On the one side, they can do a gap analysis on what is missing to improve their offering by looking at the resources already available in the Marketplace; on the other side, they can use the Marketplace to promote their services.

Across disciplines

The SSHOC project successfully bridged the silos between the SSH ESFRIs and research communities, during the 40-month lifetime of the project. Moreover, thanks to specific methodological resources such as the workflows, it is also possible to benefit from a stronger degree of cross-disciplinarity when, for instance, a method deriving from the physics domain is used in social science research to analyse specific types of data. The marketplace allows this kind of connections. Some examples on cross-disciplinary opportunities are also provided by the science projects led by SSHOC within EOSC Future.

Future developments

The SSH Open Marketplace will be upscaled to better tackle the challenge of multilinguality. Also, better connections with other services from the SSH Open Cluster or beyond, such as the Language Resource Switchboard, the Virtual Collection Registry or the GoTriple platform, are also foreseen.

Sustainability for an EOSC in practice

Three ERICs are committed to sustaining this platform financially under the SSH Open Cluster, an instrument of further collaboration between different SSH community stakeholders taking an active role in promoting the quality and impact of SSH within the European Research Area and beyond.

Future funding model scenarios

Three ERICs, DARIAH, CLARIN and CESSDA, are committed to sustain the platform until the end of 2023. The sustainability activities include an Editorial Board taking care of the platform curation and moderation, as well as participation in events where the use of the platform is promoted to interested stakeholders. The objective is to further extend the agreement under the SSH Open Cluster Memorandum of Understanding, to enhance mutual interaction, build upon and expand existing synergies and expertise, and support sharing of know-how in all areas of common interest. A sustainability plan has been developed in this regard and is available here.

Useful material related to this story

» SSHOC Service Catalogue
» SSH Open Marketplace
» Training Material
» SSHOC Zenodo Collection
» Who needs directories: a dh 2020 forum proposal

Want to learn more about the other services being developed by **SSHOC**? Read here

Liked this **#EOSCinPractice** story?
Follow @EOSCFuture for more!

Bringing big science experiment data to the researchers' fingertips.

An EOSC in Practice Story where ESCAPE's Data Infrastructure for Open Science and CS3MESH4EOSC' Science Mesh connect.

The project involved

ESCAPE is one of the five thematically clustered European Strategy Forum on Research Infrastructures (ESRFI) projects supported under the European Union H2020 research and innovation programme (Grant Agreement 824064). It aims to establish a single collaborative cluster of next generation ESFRI facilities in the area of astronomy and accelerator-based particle physics in order to implement a functional connection with EOSC. This goal is driven by the observation that sciences are facing unprecedented volumes of data and files to manage. To facilitate researchers' work, ESCAPE enables **technical interoperability between the facilities**, that is, the "ability of different information technology systems and software applications to communicate and exchange data"[1]. This minimises fragmentation, encouraging cross-fertilisation and developing joint capabilities in **astronomy, astrophysics and particle astrophysics communities**.

CS3MESH4EOSC offers Interactive and agile/responsive sharing mesh of storage, data and applications for EOSC. The project receives funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement no. 863353. CS3MESH4EOSC addresses the challenges of the fragmentation of file and application services, digital sovereignty and the application of FAIR principles in the everyday practice of researchers. Initially, 7 major data services will be combined into ScienceMesh - a **federated service mesh** providing a frictionless collaboration platform for hundreds of thousands of users (researchers, engineers, students and staff). The service mesh offers easy access to data across institutional and geographical boundaries. The infrastructure will be gradually expanded and offered to the entire education and research community in Europe and beyond.

> *"Fundamental science is at the base of our world's functioning. You need fundamental science to boost innovation, that's why we want to bring big science data at the fingertips of everyday researchers"*
>
> **Xavier Espinal, Data Infrastructure for Open Science work package leader, ESCAPE**

The Users

This EOSC in Practice story targets both researchers involved in large science projects (via ESCAPE) and citizen scientists or users interested in accessing (a part of) big science experiment data for everyday research purposes (via CS3MESH4EOSC).

The Challenge

One of the ESCAPE services is the ESCAPE DIOS (Data Infrastructure for Open Science), a scalable **federated data infrastructure** allowing an open access data service for the ESFRI projects within ESCAPE and concerned with Exabyte-scale data volumes. The **ESCAPE DIOS** is a **flexible and robust Data Lake** in terms of storage, security, safety and transfer, as well as a basic orchestration machinery, which enables the combination of technology with high quality data from different communities and, therefore, the exploration of new areas in science. While ESCAPE DIOS supports complex science research projects, it remains hardly manageable for other types of research needs. The challenge this EOSC in practice story wants to overcome is to bridge these two worlds by making complex science data available easily also to a long tale of lower scale scientific experiments with much less demanding computing needs, so to empower every possible user and enable advancement in many areas, including research, teaching, citizen science.

This is why **CS3MESH4EOSC and ESCAPE** are working together to enable the ScienceMesh activities into the ESCAPE DIOS. This will bridge diverse scientific data (regardless of its size) with researchers, outreach activities and open science initiatives.

The envisaged solution

In its research efforts, ESCAPE is slowly being adopted beyond particle physics and transitioning towards the individual researchers or citizen scientists, the kind of audience targeted by CS3MESH4EOSC. The two projects are trying to extend their boundaries and capabilities in order to eventually meet halfway. The key point is the common technology. Both projects are offering their own **Data Lakes or sync & share services** via notebooks and analysis platforms. Exploiting this common element, the aim is to enable everyone to access complex data science in a way that hides all the complexity for the benefit of the user. Enabling common login, users can access both their daily workspace and advanced science experiment data in a common blended platform. ESCAPE is currently working on a pilot case to implement the final service. The Low-Frequency Array, also known as LOFAR, can be a possible catalyser, as it is both active in ESCAPE and is bringing applied use-cases in ScienceMesh. It is a large radio telescope network where its innovation lies on combining data signals from separated antennas, by digitising them and then transported to a central digital processor, which will combine it in a software that emulates a conventional antenna. LOFAR aims to make much easier to Astronomy researchers to share and process this data.

Why do I need EOSC?

Firstly, establishing a free Data Lake and Data Management service in EOSC could help support the long-term existence of such services coming from the collaboration between ESCAPE and CS3MESH4EOSC, which would be potentially otherwise gone at the end of the projects. Moreover, Joining EOSC allows the following advantages:

» Become an integral solution easily deployable for European projects, experiments and collaborations
» Single point of access to big data from science experiments
» Easy interaction with data, all complexity is hidden
» Fast run time

The impact on society

This EOSC in practice story ensures increased inclusiveness in data access and management. Moreover, fundamental science is at the base of our world's functioning and has (although indirectly) an impact on virtually anything we deal with in daily life, from touchscreens to smartphones, internet, and nanotechnology with its related industrial applications. That's why lots of innovation can come from **sharing fundamental science experiment data**.

Contribution to cross-disciplinarity

Cross-disciplinarity is addressed on a double level in this case. First, in the aspect of pure science, via enabling data access to a multitude of different users. Second, from the technology point of view. Currently, projects and initiatives involve different techniques, systems and technologies, e.g. HPC, commercial clouds, private cloud from the universities, etc. Harmonising the way we access these technologies and the way they communicate with each other could and should boost European Research and Innovation.

Possible risks & limitations

Coordination efforts are needed to ensure the successful launch, run and maintenance of the service. The community needs to be engaged and attracted to use such service so to guarantee its future sustainability.

Sustainability for an EOSC in practice

To overcome the mentioned risks, early engagement of vast communities of final users is fundamental. Such communities could be initially unaware of their needs for data. But it is no secret that we all are indeed and will increasingly be in need for data. Let's just think about the fact that nowadays we take more pictures with our phones and need more Gigabytes in our drives. Our pictures are also larger and occupy more space than they did only a couple of years ago, because cameras are more precise and sensitive and produce better quality images. The same thing is happening with scientific instruments to measure nature: detectors, telescopes and antennas. It is not hard to imagine the computing and data processing needs many of us will face in the future. Connecting the services of ESCAPE and CS3MESH4EOSC is just a first brick that could potentially lead to many more synergies and unlock the potential of data sharing in our society.

The pilot test under development is expected to be completed in June. Regarding ESCAPE, the final aim is to have ESCAPE DIOS and all the other services produced by ESCAPE as a resource on **EOSC Exchange**. At the moment, ESCAPE services and tools are already being widely adopted by different science projects. This demonstrates the shared Data Lake infrastructure is a promising solution. And once ESCAPE's outcomes will be available via the EOSC Portal, engaging community usage will be key to the future development and further progress of the service. CS3MESH4EOSC will start working on the ScienceMesh integration into EOSC by June 2022, to be finalised by January 2023.

Future funding model scenarios

ESFRI and Research Infrastructures that will be adopting Data Lake tools will be maintaining and bearing related costs themselves to support their experiments, as the scale they operate is large.

Regarding the outputs ESCAPE is producing with CS3MESH4EOSC, the scenario is still uncertain, as the collaboration spontaneously started and was not planned at the moment of project funding. A basic infrastructure produced by ESCAPE will be continued in the framework of EOSC-Future. The rest of the services produced by ESCAPE and CS3MESH4EOSC could be transferred to EOSC when the project ends. The financial details and the possible economic support coming from EOSC itself or other interested stakeholders are yet to be decided. The collaboration with CS3MESH4EOSC, which has planned to produce a deliverable on sustainability in the final months of the project ending in December 2022, could strengthen both projects' position in the dialogue with EOSC and help drafting a comprehensive business plan.

Useful material related to this story

» Science Mesh Workshop "Global Platform for Scientific Collaboration"
» Scientific disciplines embracing no border Research Environment: ESCAPE & Science Mesh

Want to learn more about DIOS and the other ESCAPE services? Access here

Want to become a ScienceMesh adopter yourself? Access here

Liked this **#EOSCinPractice** story? Follow @EOSCFuture for more!

The Future of Seas and Oceans Flagship Initiative

Stories from projects funded under call BG-07-2019-2020

Supporting marine data discovery and accessibility to enable cross-domain research

An EOSC in Practice Story where a common interface is provided for discovery and retrieval of marine data

The project involved

Blue-Cloud
Piloting innovative services for Marine Research & the Blue Economy

Blue-Cloud is the flagship initiative of the H2020 Future of the seas and oceans programme of the European Commission (Grant Agreement no 862409). The project delivers a collaborative virtual environment to enhance FAIR and Open Science in the marine domain. Started in October 2019, Blue-Cloud is deploying a cyber platform with smart federation of an unprecedented wealth of multidisciplinary data repositories, analytical tools, and computing facilities to explore and demonstrate the potential of cloud-based Open Science and address ocean sustainability, EU Green Deal, UN Ocean Decade, and G7 Future of the Oceans objectives.

"The future prospective is to realise what is called the Digital Twin of the Oceans (DTO), this is basically the realisation of a perfect, digital, image of the real world. This will be the basis in order to execute a huge variety of "what if" scenarios and data models."
Dick Schaap, Manager at MARIS B.V. and Blue-Cloud Technical Coordinator

The Challenge

There are several research infrastructures or other data services running in Europe that cover a multitude of marine-related sciences, providing specific datasets coming from observations collected with different methods. These infrastructures constitute a diverse world, each looking at a piece of the big picture. Blue-Cloud aims to overcome fragmentation and create a bridge between thematic science clusters - such as humanities, climate, food and agriculture sciences - and EOSC, creating a data federation and providing a common access to a so-called thematic EOSC for marine data to enhance the visibility and discoverability of data from marine and environmental domains.

The Blue-Cloud Open platform is a great example of implementation of the three main components of the EOSC Federation - EOSC-Core, EOSC-Exchange and the Federation of Data & Data Services - into a unique Open Science platform accessible and usable by other research communities. Blue-Cloud is a front runner of Data federation in practice, given its success in bringing together 9 data providers covering 10M datasets from the marine domain and making them available via the DD&AS and the VRE to users, allowing interdisciplinary interactions between disciplines, thus demonstrating the value of bringing together a variety of providers and users within EOSC.

### The solution – Data Discovery and Access Service

To overcome this fragmentation, a dedicated Data Discovery and Access Service has been implemented. This service facilitates discovery and retrieval of data sets and data products for external users in stand-alone mode, also interoperable with other Blue-Cloud services, such as the Virtual Research Environment. More than 10 million data sets are managed in blue data infrastructures (BDIs) that are connected to this Blue-Cloud service to serve federated discovery and access. A common interface is provided for discovery and retrieval of data sets and data products from each of the federated BDIs. The interface also includes facilities for mapping and viewing the locations of data sets, as this is part of the query dialogue. Moreover, the interface has a shopping mechanism, enabling users to compose and submit mixed shopping baskets with requests for data sets from multiple BDIs. The service is provided on a double level. At the first level, users can identify interesting data collections in an aggregated way in a common catalogue including entries from all the federated BDIs. At the second level, users can get more specific and granular data.

Blue-Cloud currently uses **EUDAT's B2FIND** service to ingest and publish metadata referencing Blue-Cloud's data in B2FIND's interdisciplinary discovery portal. This helps make domain-specific Blue-Cloud data resources visible and discoverable on a cross-domain level. This is given even greater visibility by the fact that B2FIND is provided by the EOSC Portal as one of the generic data discovery services.
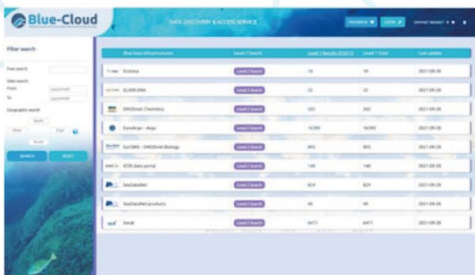
The service provider

MARIS B.V. is a private company located in Nooddorp (The Netherlands), specialised in developing web-based services and applications in the fields of oceanographic data management, geographical interfaces and websites with complex information ures.

The Users

Researchers in the marine environment can exploit Blue-Cloud services to find a broad variety of marine data and resources. More generally, researchers in other areas, such as environmental and social sciences, climate science, as well as business and industry players, can discover and use marine data for cross-domain research.

EOSC Future

EOSC in practice story #11

🏷️ **Keywords:**
#service, #data, #access, #discovery,
#infrastructure, #marine #federation
#EOSCinPractice #BlueEconomy

Why do I need EOSC?

Blue-Cloud services will bring the following benefits to their users being a provider on EOSC:

» A pilot **thematic EOSC as a role model** for the development of other thematic clouds. The cyber-platform of Blue-Cloud provides FAIR access to multidisciplinary data, analytical tools and computing and storage facilities that support research.

» **Blue-Cloud Services** showcased through five Demonstrators for oceans, seas and freshwater bodies for ecosystems research, conservation, forecasting and innovation in the Blue Economy, and making innovative use of seamless access to multidisciplinary data, algorithms, and computing resources – accelerating cross-discipline science.

» A **methodology for researchers interacting with e-infrastructure developers** to establish a cyber platform with tools and services, which support multiple scientific challenges and are fit-for-purpose, while built upon generic core principles and services.

» A **mechanism to easily access and discover blue data**. Blue-Cloud partners manage important volumes of blue data (e.g. SeaDataNet, EMODnet, CMEMS, etc.) and links have been established with major European observing networks to increase the data volume.

» APIs to access blue services that will **complement EOSC base services** providing blue thematic functionalities.

» Dynamic examples of how a framework like Blue-Cloud can **address one or several of the policy challenges** defined in the Bioeconomy Strategy, the Circular Economy Strategy, the Blue Growth Strategy, the Common Fisheries Policy, the Maritime Spatial Planning Directive, the Marine Strategy Framework Directive, the International Ocean Governance Communication and the UN SDGs.

» **A global Blue Economy community** close to the EOSC vision, including the marine and maritime industry.

» The opportunity of **bringing EOSC in the Blue Economy long-term vision** via the policy-oriented Blue-Cloud Roadmap to 2030 which seeks a series of EU Calls for further development and uptake of the Blue Cloud by multiple VRE applications and connecting additional marine data infrastructures.

Societal impact

The Marine Strategy Framework Directive, the European strategy for green management governance, the Sustainable Development Goals set up by the United Nations and obviously the EU Green Deal represent major political measures whose implementation is supported also by the enhanced awareness and sensitivity that initiatives like Blue-Cloud have on the socio-political European and international environment. BlueCloud's role is also to inform interested stakeholders that in order to implement these initiatives and understand if a sustainable practice is feasible, scientists need to study its economical and social impacts. Therefore, a multidisciplinary approach is fundamental to plan future business practices.

Across disciplines

The collaboration between EOSC and Blue-Cloud is thus an effective way to foster the adoption of a multidisciplinary approach. In this particular case, interaction among marine and socio economic domains is critical to understand their feasibility and impact.

Future developments

Blue-Cloud aims to receive more feedback from the user base about the services developed. A first step is the Strategic Roadmap to 2030, for which interviews were conducted with key stakeholders to understand if the services developed indeed meet the user's needs. A wide variety of dissemination and promotional activities have been planned and carried out to increase the visibility of Blue-Cloud services and

stimulate their uptake, such as: articles, conferences, workshops, joint workshops and a highly successful Hackathon. Widening the user base is fundamental for feedback collection, and provides developers with important information to refine the Blue-Cloud services.
In addition, Blue-Cloud is set to improve quality of data through the validation of data sets so that the data models can run better.

Sustainability for an EOSC in practice

Copernicus Marine Service and EMODnet are the two main Research Infrastructures (RI) that are funded and sustained by the European Commission. Blue-Cloud is federating data from both Copernicus Marine Service and EMODnet RIs, aiming to create an environment with shared resources from multiple domains.

Future funding model scenarios

Blue-Cloud supports the creation of the Digital Twin of the Ocean (DTO), the realisation of a digital image of the real ocean. The DTO is also part of the Digital Ocean Knowledge System under the EU Mission Restore our Ocean and Waters. The way forward is to build a business plan to detail options and scenarios on how the strong network of Blue-Cloud stakeholders could ingest and exploit the results of the initiative. The recently awarded Blue-Cloud 2026 project (January 2023 – June 2026) will further evolve this pilot ecosystem into a **Federated European Ecosystem to deliver FAIR & Open data and analytical services**, instrumental for deepening research of oceans, EU seas, coastal & inland waters. Blue-Cloud 2026 will expand the federated approach of Blue-Cloud, involving more aquatic data stakeholders, and interacting with EOSC developments, in support of the EU Green Deal, UN SDG, EU Destination Earth, and the EU Mission Starfish on healthy oceans, seas, coastal and inland waters, mobilising and making available major additional data resources as validated and harmonised in-situ data by means of Data Lakes.

Useful material related to this story

» Blue-Cloud Data Discovery & Access Service
» Blue-Cloud Hackathon pilots

Want to learn more about the other services being developed by **Blue-Cloud**? Read here

Liked this **#EOSCinPractice** story? Follow @EOSCFuture for more!

EOSC Future

Providing computing platforms and analytical services to facilitate the collaboration between researchers

An EOSC in Practice Story where a common workspace is provided to researchers in the marine world to analyse, process, share data and co-develop research products

The project involved

Blue-Cloud is the flagship initiative of the H2020 Future of the seas and oceans programme of the European Commission (Grant Agreement no 862409). The project delivers a collaborative virtual environment to enhance FAIR and Open Science in the marine domain. Started in October 2019, Blue-Cloud is deploying a cyber platform with smart federation of an unprecedented wealth of multidisciplinary data repositories, analytical tools, and computing facilities to explore and demonstrate the potential of cloud-based Open Science and address ocean sustainability, EU Green Deal, UN Ocean Decade, and G7 Future of the Oceans objectives. In January 2026, a follow-up project, **Blue-Cloud 2026**, has been funded to further expand the ecosystem and the core services offered to researchers of oceans, seas, coastal and inland waters.

The Challenge

There are several research infrastructures or other data services running in Europe that cover a multitude of marine-related sciences, providing specific datasets coming from observations collected with different methods. These infrastructures constitute a diverse world, each looking at a piece of the big picture, sometimes hindering collaboration and data sharing. Blue-Cloud aims to overcome fragmentation and build a bridge between thematic science clusters - such as marine, climate, food and agriculture sciences - and EOSC, creating a data federation and providing a common access to a so-called thematic EOSC for marine data. By connecting leading marine data management infrastructures with horizontal e-infrastructures, the project aims to maximise the exploitation of data resources available from different sources. The Blue-Cloud framework consists of two major technical components:

(1) a Blue-Cloud Data Discovery and Access service, already presented in a previous EOSC in practice story, to serve federated discovery and access to blue data infrastructures, and

(2) a Blue-Cloud Virtual Research Environment (VRE) to provide computing platforms and analytical services facilitating the collaboration between researchers, which is detailed hereafter.

> *"Blue-Cloud VRE, backed by the D4Science infrastructure, enriches EOSC with a community of 18k potential users and new services for researchers which are available in the EOSC Marketplace."*
>
> **Pasquale Pagano, Senior Researcher at the InfraScience Laboratory @ Italian National Research Council (CNR)**

The solution – Virtual Research Environment

The Blue-Cloud Virtual Research Environment (VRE) facilitates collaborative research using a variety of data sets, analytical tools, and computing services that allow data access, processing, harmonisation, sharing, publishing, as well as the (co-)creation of new research products that can be stored in the same environment and shared with other users. Within the Blue-Cloud VRE, a set of analytical workflows facilitate the creation of shared workspaces that enable open science, relying on data sources and inputs that can be retrieved from the blue data infrastructures by means of the Blue-Cloud Data Discovery and Access, or can be retrieved from different other data portals and resources. The key value of the VRE is therefore to allow scientists and practitioners not only to easily access different sets of marine data but also to process and experiment with them via the analytical and visual tools made available by each demonstrator. The various VRE functionalities can be accessed logging into the Blue-Cloud Gateway via Single Sign On (SSO) and can be used in different ways based on the user needs.

For instance, if we consider data processing and analytics, users may exploit different types of support as shown in the examples below:

» **Use of standard technologies**, such as notebooks (The Blue-Cloud VRE has integrated **JupyterHub**) to create an environment and develop codes in any programming languages. The notebook is connected to the storage workspace of the infrastructure of the Blue-Cloud VRE and enables access to the data collected via the DD&AS. The notebook can be shared with other users to cooperate with collaborators

» A more advanced approach would be to use tools that allow **combining specific programming languages**. For instance, researchers could start a process in a notebook elaborating some data with a code in Python, but then continue the data analysis in R, which is one of the most used languages in the marine research domain, relying on a specific environment available in Blue-Cloud VRE, namely **RStudio**.

» An even more sophisticated type of usage is **high-throughput computing** (HTC) which allows elaborating large quantities of data relying on distributed computing facilities. By developing a code in any programming languages and registering it to the VRE, the analysis will be executed in a series of machines in the cloud, which are assigned to the research group and the result of the data processing will be available to all the members of the storage workspace

The Blue-Cloud VRE hosts to date **5 specific demonstrators** developing Virtual Labs for the benefit of the wider community, in order to demonstrate Blue-Cloud's potential in different fields of marine research, ranging from biodiversity to environmental science, as well as fisheries and aquaculture. An additional Lab, named Blue-Cloud Lab, offers access to common services as open-access facilities.

Overview of Blue-Cloud VRE in the Blue-Cloud Gateway
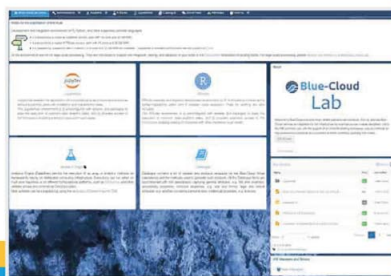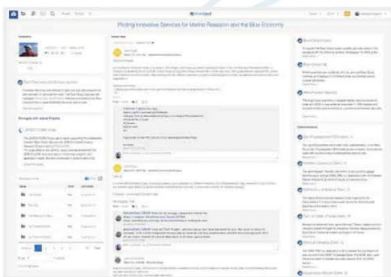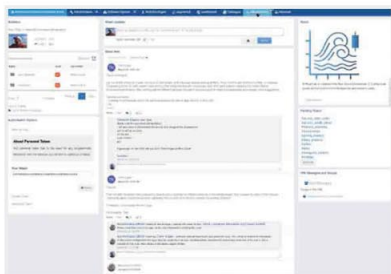
The service provider

The Blue-Cloud VRE is developed by the Italian National Research Council (CNR), based upon the existing D4Science e-infrastructure as developed and managed by CNR-ISTI. This e-infrastructure already hosts multiple Virtual Labs and offers a variety of services, which can be adopted for Blue-Cloud. The D4Science e-infrastructure also has proven solutions for connecting to external computing platforms and means for orchestrating distributed services, which will be instrumental for smart connections to the other e-infrastructures in the Blue-Cloud ecosystem.

The Users

Researchers and Scientists in the marine environment can exploit Blue-Cloud services to find a broad variety of marine data and resources. More generally, researchers in other areas, such as environmental and social sciences, climate science, as well as business and industry players, can discover and use marine data for cross-domain research.

What the users say

"Thanks to the RStudio component offered by the Blue-Cloud VRE, I was able to reduce the computing time from hours to minutes when running the model that quantifies the relative contributions of the drivers in phytoplankton dynamics, as part of the Zoo and Phytoplankton EOV Demonstrator".

Viviana Otero (Flanders Marine Institute - VLIZ)

Why do I need EOSC?

Blue-Cloud services will bring the following benefits to their users being a provider on EOSC:

» A pilot **thematic EOSC as a role model** for the development of other thematic clouds. The cyber-platform of Blue-Cloud provides FAIR access to multidisciplinary data, analytical tools and computing and storage facilities that support research.

» **Blue-Cloud Services** showcased through five Demonstrators for oceans, seas and freshwater bodies for ecosystems research, conservation, forecasting and innovation in the Blue Economy, and making innovative use of seamless access to multidisciplinary data, algorithms, and computing resources - accelerating cross-discipline science.

» A **methodology for researchers interacting with e-infrastructure developers** to establish a cyber platform with tools and services, which support multiple scientific challenges and are fit-for-purpose, while built upon generic core principles and services.

» A **mechanism to easily access and discover blue data**. Blue-Cloud partners manage important volumes of blue data (e.g. SeaDataNet, EMODnet, CMEMS, etc.) and links have been established with major European observing networks to increase the data volume.

» APIs to access blue services that will **complement EOSC base services** providing blue thematic functionalities.

» Dynamic examples of how a framework like Blue-Cloud can **address one or several of the policy challenges** defined in the Bioeconomy Strategy, the Circular Economy Strategy, the Blue Growth Strategy, the Common Fisheries Policy, the Maritime Spatial Planning Directive, the Marine Strategy Framework Directive, the International Ocean Governance Communication and the UN SDGs.

- » **A global Blue Economy community** close to the EOSC vision, including the marine and maritime industry.
- » The opportunity of **bringing EOSC in the Blue Economy long-term vision** via the policy-oriented Blue-Cloud Roadmap to 2030 which seeks a series of EU Calls for further development and uptake of the Blue Cloud by multiple VRE applications and connecting additional marine data infrastructures.

Societal impact

The Marine Strategy Framework Directive, the European strategy for green management governance, the Sustainable Development Goals set up by the United Nations and obviously the EU Green Deal represent major political measures whose implementation is supported also by the enhanced awareness and sensitivity that initiatives like Blue-Cloud have on the socio-political European and international environment. BlueCloud's role is also to inform interested stakeholders that in order to implement these initiatives and understand if a sustainable practice is feasible, scientists need to study its economical and social impacts. Therefore, a multidisciplinary approach is fundamental to plan future business practices.

Across disciplines

The collaboration between EOSC and Blue-Cloud is thus an effective way to foster the adoption of a multidisciplinary approach. Many of the teams involved in the Blue-Cloud VRE already require professionals with different profiles and generate metrics and indicators that combine multiple perspectives. In this particular case, the interaction among marine and socio economic domains is critical to understand their feasibility and impact.

Future developments

Blue-Cloud aims to receive more feedback from the user base about the services developed. A first step is the Strategic Roadmap to 2030, for which interviews were conducted with key stakeholders to understand if the services developed indeed meet the user's needs. A wide variety of dissemination and promotional activities have been planned and carried out to increase the visibility of Blue-Cloud services and stimulate their uptake, such as: articles, conferences, workshops, joint workshops and a highly successful Hackathon. Widening the user base is fundamental for feedback collection, and provides developers with important information to refine the Blue-Cloud services.

In addition, Blue-Cloud is set to improve quality of data through the validation of data sets so that the data models can run better.

Sustainability for an EOSC in practice

Copernicus Marine Service and EMODnet are the two main Research Infrastructures (RI) that are funded and sustained by the European Commission. Blue-Cloud is federating data from both Copernicus Marine Service and EMODnet RIs, aiming to create an environment with shared resources from multiple domains. Blue-Cloud services are also supported by D4Science which has a self-sustained infrastructure listed among the strategic infrastructural assets and resources of national interest in Italy.

Future funding model scenarios

Blue-Cloud supports the creation of the Digital Twin of the Ocean (DTO), the realisation of a digital image of the real ocean. The DTO is also part of the Digital Ocean Knowledge System under the EU Mission Restore our Ocean and Waters. The way forward is to build a business plan to detail options and scenarios on how the strong network of Blue-Cloud stakeholders could ingest and exploit the results of the initiative. The recently awarded Blue-Cloud 2026 project (January 2023 - June 2026) will further evolve this pilot ecosystem into a **Federated European Ecosystem to deliver FAIR & Open data and analytical services**, instrumental for deepening research of oceans, EU seas, coastal & inland waters. Blue-Cloud 2026 will expand the federated approach of Blue-Cloud, involving more aquatic data stakeholders, and interacting with EOSC developments, in support of the EU Green Deal, UN SDG, EU Destination Earth, and the EU Mission Starfish on healthy oceans, seas, coastal and inland waters, mobilising and making available major additional data resources as validated and harmonised in-situ data by means of Data Lakes.
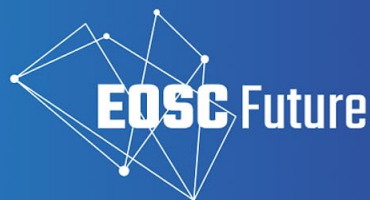
Useful material related to this story

- » Blue-Cloud Gateway @D4Science
- » Blue-Cloud VRE
- » Blue-Cloud Data Discovery & Access Service
- » Blue-Cloud Hackathon pilots
- » Blue-Cloud Services on the EOSC Marketplace

Want to learn more about the other services being developed by **Blue-Cloud**? Read here
Liked this **#EOSCinPractice** story?
Follow @EOSCFuture for more!

EOSC Future

Liked these #EOSCinPractice stories?

Follow @EOSCFuture for more!